

UNIVERSIDAD DE ALCALÁ
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN, ESCUELA POLITÉCNICA



TESIS DOCTORAL

**VISUALIZACIÓN DE ESQUEMAS DE REPRESENTACIÓN DE
CONOCIMIENTO PARA EL ACCESO A RECURSOS EN REPOSITARIOS
DIGITALES**

Autor:

Paulo Alonso Gaona García

Directores:

Dr. Salvador Sánchez Alonso

Doctor en Informática

Universidad de Alcalá

Dra. Ana Feroso García

Doctora en Informática

Universidad Pontificia de Salamanca

Alcalá de Henares, Mayo 9 de 2014

UNIVERSITY OF ALCALÁ
DEPARTMENT OF COMPUTER SCIENCE, POLYTECHNIC SCHOOL



DOCTORAL THESIS

**VISUALIZATION OF KNOWLEDGE REPRESENTATION SCHEMES TO
ACCESS RESOURCES IN DIGITAL REPOSITORIES**

Author:

Paulo Alonso Gaona García

Supervisors:

PhD. Salvador Sánchez Alonso

University of Alcalá

PhD. Ana Fermoso García

Pontifical University of Salamanca

Alcalá de Henares, May 9 2014

A Dios por su amor y fidelidad,
A Margarita por todo lo que has sido para mi vida,
A mi familia por el tiempo de espera y apoyo incondicional

*"Lo que sabemos es una gota,
lo que ignoramos un inmenso océano.
La admirable disposición y armonía
del universo, no ha podido sino salir del
plan de un Ser omnisciente y omnipotente"*

(Isaac Newton)

Agradecimientos

En primera instancia quiero agradecer especialmente a Dios, por bendecirme y permitir que este proyecto fuera toda una realidad. Por su sustento, cuidado y respaldo en todas las circunstancias que surgieron a lo largo de esta travesía; por levantarme siempre en los momentos de dificultad. A él sea toda la gloria y toda la honra.

Ha sido todo un privilegio el poder contar con el acompañamiento y seguimiento de dos personas maravillosas en la dirección de este proyecto, no solo por sus valiosos aportes y apoyo incondicional; sino por el compromiso, dedicación, y todo el proceso de formación que transmitieron a lo largo de toda esta ruta de navegación. Quiero agradecer a mi director el Dr. Salvador Sánchez Alonso, por abrir la puerta para emprender este viaje, que como bien lo señalaba en una primera ocasión, el propósito de la tesis es compromiso al 100% para llegar a buen puerto. Gracias a su dirección, seriedad y compromiso, fue posible que este viaje cumpliera su propósito, sin importar las dificultades o aquellos periodos donde no se presentaba buen viento para continuar con el recorrido. No puedo dejar de expresar mi agradecimiento a mi co-directora, la Dra. Ana Feroso García; por su dirección y disposición que fueron de aliento para seguir por el camino en el desarrollo y culminación de este proyecto. A ellos, mis más sinceros agradecimientos, porque más allá de un proceso de formación, se ha forjado una gran amistad.

De la misma forma quisiera agradecer a todos mis colegas, compañeros y amigos del grupo de investigación IERU. Al Dr. Miguel-Ángel Sicilia por sus acertados consejos, a la Dr. Elena García Barriocanal por facilitar las actividades de trabajo dentro del grupo de investigación. Un agradecimiento especial para David Moncunill, quien de manera oportuna dio un gran empujón a este proyecto aportando grandes ideas en el área de usabilidad. Una mención especial para mis compañeros y amigos de formación doctoral, Karina, Enayat, Rutilio y Bernabé, quienes de manera cronológica llegaron al grupo en los momentos más apropiados para ser de gran ayuda a través de sus aportes y constantes palabras de aliento dentro de mi proceso de formación como persona e investigador; para ellos mi mensaje de ánimo y aliento en cada uno de sus proyectos que están emprendiendo y a sus familias. No podía olvidar mencionar a los demás compañeros y amigos del grupo IERU: Eydel, Phillip, Alberto, Meritxell, Juan Moreno, David Baños, Jaakko, Paco, Juan Pablo, Abel y Miguel Refusta por la ayuda incondicional en el proyecto, por sus aportes y de manera general por facilitar una estancia amena en el laboratorio, mis mejores deseos para todos.

Al grupo de personas que facilitaron el desarrollo de mis actividades de investigación en la estancia doctoral realizada en Atenas-Grecia. Al Dr. Giannis Stoitsis por todo el acompañamiento, seguimiento y aportes realizados dentro de las actividades de trabajo realizados en Atenas, a Nikos Mannolis y Vasilis Protonotarios por su apoyo en el área técnica y acceso a servicios en el proyecto, y de manera especial a todo el grupo de trabajo de Agro-

Know en Atenas, fue una grata experiencia el poder contar y compartir con cada uno de ustedes. Una mención especial para Paco y Effie por sus revisiones y observaciones en varias de las publicaciones realizadas en este proyecto.

También agradecer a la Universidad Distrital Francisco José de Caldas por todo el apoyo que me dio para llevar a cabo este proceso de formación doctoral en la Universidad de Alcalá. A todas las personas que colaboraron y facilitaron los procesos y tramitaciones administrativas para que este proyecto fuera una realidad. A todos y cada uno de mis amigos, colegas y compañeros de trabajo de la Universidad Distrital: a Carlos, Julio, Sandro, Giovanny, Luz, Herminia, Susana, Beatriz, Javier y Víctor; quienes formaron parte de este proceso en los momentos más indicados. Al grupo de investigación GIIIRA, en especial a Diego y Carlos, y cada uno de sus integrantes que estuvieron pendientes desde Colombia de todo este proceso, aportando ideas y conocimiento.

No podría dejar de lado el acompañamiento incondicional y fundamental de mis familiares más cercanos. A mi esposa Margarita, por acompañarme en esta gran aventura, por el ánimo constante, palabras de aliento y creer siempre en mis capacidades. A mis padres: Doris y Martin por todo su esfuerzo y sacrificio en darnos siempre lo mejor para nuestra formación como personas y ser ese sustento y pilar para la familia, por su ejemplo de constancia y trabajo para que todo esto fuera una realidad. A mi hermano Elvis, por ser siempre ese ejemplo de formación para mi vida. A mi hermana Ángela por cada mensaje de aliento y ánimo, a mis sobrinos: Juan Diego, Juan José, Catalina y Martin, porque a pesar de su corta edad, serán el pilar de muchos éxitos y alegrías para nuestras vidas el día de mañana, junto con todos los nuevos integrantes que lleguen a nuestra familia. A los tíos: Tatiana, María Fernanda, Hans y Farid, gracias por el cuidado, servicio y atención que tuvieron siempre a la distancia; al igual que todos los aquí citados, y muchas personas que por olvido o descuido no menciono aquí, les estaré siempre agradecido.

Resumen

El siguiente documento presenta los resultados de investigación realizados a partir de estudios enfocados en el desarrollo e implementación de interfaces de búsqueda de objetos de aprendizaje, a partir de técnicas de visualización sobre repositorios digitales. Actualmente existen una gran cantidad de recursos digitales sobre Internet, y el acceso a los mismos en gran medida dependen de las estrategias que puedan ofrecer motores de búsquedas convencionales, o soluciones especializadas que permitan su clasificación, gestión y administración, como es el caso de los repositorios digitales. Sin embargo, existen una serie de factores que influyen sobre el acceso a los mismos, partiendo de la definición de los metadatos, y las estrategias de búsqueda que se definan sobre grandes volúmenes de información. Una de las áreas de mayor aceptación a lo largo de los últimos años es la visualización de información, área de trabajo que facilita la presentación visual de información compleja haciendo uso adecuado de espacios y estructuras gráficas, con el fin de facilitar su rápida asimilación y comprensión. Por lo tanto, para los propósitos específicos de esta investigación, abordaremos el área de visualización de información mediante el uso de metodologías de evaluación y estrategias de diseño para el desarrollo e implementación de interfaces de búsquedas efectivas, para el acceso a colecciones de recursos digitales alojados en repositorios digitales. El propósito fundamental de esta investigación es ofrecer alternativas de acceso a partir de técnicas de visualización, para facilitar a creadores de repositorios digitales el análisis, desarrollo e implementación de interfaces de búsqueda visual.

Palabras Clave

Visualización de información, técnicas de visualización, esquemas de representación de conocimiento, repositorios digitales, objetos de aprendizaje, metadatos, calidad de metadatos, usabilidad.

Abstract

The following document presents the results of research conducted from studies focused on the development and implementation of search interfaces in learning objects, based on visualization techniques over digital repositories. Currently there are a lot of digital resources on Internet and access to them to a large extent depend on the strategies that can offer conventional search engines or specialized solutions that allow their classification, management and administration, as is the case of digital repositories. However, there are a number of factors that influence access to digital resources, based on the definition of metadata, and search strategies based on large volumes of information. One area in greater acceptance over the last few years is the information visualization workspace, area that facilitates the visual presentation of complex information spaces using appropriate graphs and structures in order to facilitate their rapid assimilation and understanding. Therefore, for the specific purposes of this research, we will address the information visualization area by using assessment methodologies and design strategies for the development and implementation of effective search interfaces for access to collections of digital resources hosted on digital repositories. The main aim of this research is to provide an alternative access based on visualization techniques to facilitate digital repositories creators, the analysis, development and implementation of visual search interfaces.

Keywords

Digital repositories, knowledge representation schemes, learning objects, metadata, metadata quality, usability, visualization information, visualization techniques.

Executive Summary

MOTIVATION

This summary presents the results of research on the development and implementation of search interfaces for learning objects that are based on visualization techniques used for digital repositories, as well as the background of this research. Currently there are thousands, or even millions, of digital resources published on the Internet, stored in digital libraries, repositories and documentation centres. However, accessibility to digital resources largely depends on the strategies offered by each traditional search engine or other specialized solutions. These solutions allow the classification, management and administration of resources, as in the case of digital repositories. Moreover, metadata is another factor that influences information access and search strategies used in the case of large amounts of information.

A widely accepted field of study during the last few years is that of information visualization workspaces that facilitate the visual representation of complex information through the use of adequate graphs and structures, that facilitate their understanding. In this research, we focus on information visualization using assessment methods and designing strategies for the development and implementation of effective search interfaces that would be able to provide access to large collections of digital resources hosted in digital repositories. The motivation of this research is to analyse various types of visualization techniques to provide a set of recommendations that would help repository creators, to analyse, develop and implement visual search interfaces.

In recent years, digital repositories have rapidly developed due to the exponential increase of the number of digital resources published (Margaryan & Littlejohn, 2008). This growth of digital resources has led to the increase in related efforts, namely: i) (in terms of technology) the development of distributed repositories, heterogeneous repositories, and federations of repositories (central access points to repositories) (McGreal, 2008), ii) (in terms of semantics) the development of linked knowledge classification schemes through the use of ontologies and thesauri to provide a better understanding and organization of digital resources (Soto, Gordo, & Sanchez-Alonso, 2007), and finally iii) (in terms of access) the definition and inclusion of metadata descriptions to provide more relevant search results (Cechinel, Sánchez, & Sicilia, 2009).

Given these strategies, then we present the main objectives of the research.

OBJECTIVES

The aim of this research is to help creators of digital repositories to provide better search services, through the use of visualization techniques, so that users can: i) find materials in an extensive collection of digital resources effectively and precisely, ii) search resources by topic or knowledge area, iii) find out whether a repository covers a certain topic according to the number

of digital resources a certain term is found in (thematic coverage) and display through a graphical structure this information, and finally, iv) use effective interfaces to browse and search for digital resources.

In the next section, we describe the most important fields and knowledge areas that were considered relevant to our research, in order to identify its background.

BACKGROUND

a. Digital repositories

Repositories provide their users with several access alternatives which are in some cases visual, providing user interfaces that allow users to perform search processes. However, previous studies have found that these interfaces do not completely meet users' needs (Tenopir, 2003). Users may have difficulty in finding resources (Khoo, Kusunoki, & MacDonald, 2012), and/or the resources found are not always relevant to their search (Nash, 2005). Moreover, navigation problems may occur when users try to return to a previously accessed record (Jeng, 2005), beside limitations related to the optimal combination of navigation and search methods (Hartson, Shivakumar, & Pérez, 2004) that do not allow displaying (in at-a-glance way) the materials available in the repository, as far as the thematic coverage is concerned (Hitchcock et al., 2003; Tsakonas & Papatheodorou, 2008). Therefore, it is difficult to judge whether it is worth searching for materials in the repository, or if it is more convenient and effective to use some other external search strategy (e.g. using general purpose search engines like Google). On the other hand, the most representative learning objects repositories in Europe provide, at least, a visual representation scheme for searching digital resources. For example: i) MERLOT^[1], using textual categories and textual search, ii) MACE^[2], using semantic, social and contextual search, iii) ARIADNE^[3], using textual categories, iv) E-LERA^[4], using textual search, v) E-LIS^[5], using textual category and textual search, and finally, vi) Organic-Edunet^[6], using semantic categories and textual search. However, access to them is not always obvious and additional information is required if the user wants to fully understand the search methods they provide.

The use of visual search methods has emerged only recently in the field of digital repositories. Some methods focus on the adequate level of access to digital resources. The MACE project (Stefaner et al., 2007) proposes several alternatives of visual search (semantic, social and contextual) for accessing digital resources in the field of "design and architecture" through classification strategies involving: keywords, location, competence, social area and facets (Stefaner & Muller, 2007). These studies focus on perspectives that use various navigation paths combined with social labelling. Studies on the use of MACE showed that the principles of multifaceted navigation facilitate immersion processes through activities of collaborative tagging (Stefaner et al., 2009); and, that the definition of metadata is crucial for the improvement of search processes based on contextual-search strategies.

Several authors (Aula & Käki, 2005; Buttenfield, 1999; Hargittai, 2004; Hearst, 2009) argue that one of the fundamental principles for building search interfaces is simplicity. For that reason, it is very important to analyze an interface from the user's point of view to maximize user satisfaction, which is the main aim of Human Computer Interface (HCI). Therefore, usability strategies are an important factor within the HCI for the development and evolution of interfaces found both in libraries and digital repositories (Buttenfield, 1999; Fox et al., 1999; White & Roth, 2009). Other studies (Marchionini, 2006; White & Roth, 2009) specifically focus on searching and browsing capabilities for locating digital resources through visual interfaces, and conclude that traditional search and browsing strategies are not sufficiently robust or flexible to facilitate location and access to a collection of digital resources. In terms of efficiency, the results of those studies demonstrate that visual interfaces often cause the underutilization of metadata information, far below their full semantic potential, even though these visual interfaces are key for the enrichment of the processes used for finding digital resources of a specific knowledge area (Cechinel, et al., 2009; Fernández, 2001).

With the rapid growth of digital resources collections in recent years, many institutions have developed a variety of strategies and policies to facilitate and encourage their use (Liew & Foo, 1999; Prabha, 2007, S. Zhang, Shen, & Ghenniwa, 2004). In the case of the educational sector, this growth has become an essential part of the learning environment (Armstrong et al., 2001; Hayati & Jowkar, 2008, Nicholas, Huntington, & Jamali, 2007; Pancheshnikov, 2007). However, there is a widespread view that this type of education-oriented resources is not used to its full potential from the point of view of improving teaching and research processes (Edward A Fox & Logan, 2005; Jain & Babbar, 2006; Shuling, 2007). This inefficacy of repositories is attributed to a number of factors associated with the lack of: i) competence, ii) training, and iii) time devoted to their use (Roberts, 1995; Shuler, 2007; Rehman & Ramzy, 2004). Moreover, big efforts and suggestions have been made for the further development of strategies for classifying and organizing digital resources based on knowledge representation schemes (thesauri and ontologies). The implementation of these strategies has been of great benefit to experts that work on activities related to the management and administration of digital repositories (creators and administrators). However, these efforts do not sufficiently focus on the final user's needs to locate and access digital resources (students, teachers).

b. Repositories search strategy

In general, digital repositories search systems provide at least two search alternatives through their user interface: simple (also known as basic search) and advanced search. The basic search does not require from the user a deep knowledge of the system and/or the search process, as it allows users to perform a quick search, but restrict them to use keywords and precise queries. However, it is considered easy to understand and use by users without prior experience. On the other hand, users need more knowledge and skills to use the advanced method. Basic searches are the most popular, while the advanced search requires more understanding. In this context, the design of an interface is of a paramount importance as it could greatly facilitate the search process and could significantly improve user satisfaction (Shneiderman & Plaisant, 2005).

c. Metadata as search criteria

Successful digital resources finding depends largely on metadata quality, that is an essential factor for obtaining relevant search results (Prieta & Gil, 2010; Muñoz, Calvillo, Ochoa, Santaolaya, & Alvarez, 2010). On the other hand, the quality of metadata is important for the improvement of the strategies used for the indexing of learning objects stored in repositories (Ochoa et al., 2005; Stuckenschmidt, et al., 2004; Wiley, 2002a). Therefore, metadata omission negatively affects the results of a search when the search criteria are predefined, for example restricted to a specific knowledge area.

Content quality is an indicator that allows the evaluation of digital resources. In this field, there are several studies presenting methods for assessing the quality of digital resources (Chuanjun, 2004; Downes, 2007; Gonçalves, Moreira, Fox, & Watson, 2007; Sanz, et al. 2009), as well as the quality of the content of collections of digital resources (Chuanjun, 2004; Downes, 2007). These methods offer a number of criteria for metadata quality assessment of a collection of digital resources. However, some authors such as Sicilia, Sánchez, & Benito (2006) indicate that there is not a global model for metadata quality criteria that can be applied to all work scenarios, thus it is necessary to establish different evaluation methods for each specific scenario.

d. Knowledge representation schemes

The development of repositories is one of the most widespread initiatives for the centralization of search processes in digital resources. Repositories are important infrastructures that facilitate access to collections of digital resources created by museums, libraries, archives and digital documentation centres. These institutions focus on activities related to the storage of educational materials in order to optimize the management and search processes for their resources. However, the large number of digital resources has generated some limitations, which in some cases also affect the use of repositories for accessing educational materials. This type of deficiencies has led to the development and use of alternative solutions based on the implementation of markup languages (semantic languages) and knowledge representation schemes that facilitate activities of classification, categorization, linking and management of contents. The implementation of these strategies has led to a solid technological structure, linked to a series of semantic enrichment strategies, based on the use of knowledge representation schemes. These solutions allow certain management and administration activities to creators and developers of repositories. Nevertheless, for a conventional user (e.g., students or teachers), the use of such educational repositories is not always easy because their interfaces and conventional search mechanisms do not always offer adequate search strategies. This is a key factor, which can even hamper learning processes, and consequently foster the gradual abandonment of these tools.

The evolution of digital repositories has allowed management centralization and provided access to thousands of digital resources. However, the rapid development of educational materials has posed a challenge to the creators of digital repositories that aim to organize, classify and manage contents. Therefore, over the last years, one of the most widely accepted

strategies used to address these needs is the use of classification tools based on knowledge representation schemes.

Literature on this topic offers a diverse list of definitions associated with the notion of knowledge representation scheme, from the perspective of philosophy, human reasoning, cognitive science and artificial intelligence. However, for the purposes of our research, we have defined each of the notions involved in the concept of a knowledge representation scheme in order to understand its value and importance in the field of digital repositories. According to Durkin (1994) knowledge is related to a subject area and to the concepts linked and associated to it. This approach allows us to identify two elements, namely: i) the thematic area, which we understand as a domain or set of topics and concepts of interest, and secondly ii) the relations that maintain certain areas of interest according to thematic, hierarchical, historical or other type of affinity in each case.

Artificial intelligence provides a number of strategies and types of knowledge representation (Russell, Norvig, & Davis, 2010). These are related to methods and structures used to encode knowledge in intelligent systems, and determine the relations of their data structure. Thus, a representation scheme includes data structures defined by the use of tables, trees, links, graphs, etc. Each data structure has advantages and disadvantages when it comes to the representation of different types of knowledge, therefore, as Gašević (2009) indicated, there is no generic method for knowledge representation that could be a standard method for structuring data in all cases. In order to have an overview of these knowledge representation structures, each user navigates through the interface, as an essential process in the field of information visualization (Graham, Kennedy, & Benyon, 2000). Thus, this concept is associated with the manner in which a user proceeds to browse through a hierarchical structure using a taxonomic representation tool.

Most computer applications use some form of knowledge associated with data. As a result, handling, administration and management of contents implies a significant workload, both operationally and from the decision-making point of view. In some cases, knowledge is represented as raw data, often stored as complex structures according to logical sequences, rules, trees, semantic graphs, and other forms of representation (Flouris, Plexousakis, & Antoniou, 2003). These structures correspond to a formal classification, and in the best case, to semantic relations of the data, at the level of dependence, association, affinity, etc.

Each knowledge representation technique requires a notation, which determines aspects related to the subject area, level of affinity, links and forms of association, among others. In order to be effective within a learning environment, a knowledge representation scheme must be consistent and as realistic as possible to represent a subject area, and its connexions to other subject areas (Chrysafiadi & Virvou, 2013). This allows, among other things, expanding or limiting the knowledge offered to the users of the scheme. The tools used for knowledge representation when designing a digital repository, are most often taxonomies, ontologies, thesauri, graphs and mind maps among others. These tools could facilitate search processes related to a specific topic or knowledge area; however, a search for learning objects could return

a huge list of results if definitions of essential elements that help the refinement of the search are not provided. Therefore, if indicators to help the quality assessment of the recorded information are not provided, searching for learning objects could prove to be an activity that implies a considerable waste of time and effort (Kumar, Nesbit, & Han, 2005).

e. Visual interfaces

In addition to languages and knowledge representation schemes, to carry out a search process, some repositories also provide alternative access through visual interfaces. Previous studies have found that some search interfaces do not completely meet the needs of the final users (Tenopir, 2003). Often, the results displayed are not relevant to the user-defined criteria (Nash, 2005). Moreover, navigation problems may occur when users try to return to previously consulted records (Jeng, 2005). Regarding this, Kim (2008) found a number of problems associated with the design of the interface of an institutional repository in Korea. The study showed that the topics of interest queried by users were not sufficiently visible, because the navigation menu was too small and too dark. Other studies have shown limitations in combining navigation strategies and search methods (Hartson, Shivakumar, & Perez, 2004), when interfaces do not allow the visualization and display by topic of materials available in a repository (in an at-first-glance way) (Hitchcock et al., 2003; Tsakonas & Papatheodorou, 2008). This last limitation is a determining factor for users to decide whether they continue using these tools or not, given that they find difficult to determine whether it is worth continuing the process of searching materials in the repository, or if they should better use some other external search strategy.

Usability studies have been made to the above-mentioned analyses on access interfaces to repositories. Those studies revealed some limitations, as we have just already mentioned. In short, these usability studies have identified a series of limitations that some libraries and digital repositories present. Tsakonas & Papatheodorou (2007) conducted a usability analysis on the digital library E-LIS, finding that it is difficult to perform certain search processes, because of it is excessively time-consuming to carry out certain searching tasks and the great effort is needed to understand the user interface. The authors highlighted the importance of complementing these strategies, given that they are not sufficiently robust and flexible on their own, so as to facilitate location and access to a collection of digital resources. Along the same lines, we should also mention the studies of Buchanan & Salako, (2009) and Petrelli (2008) that concluded that search processes based on a limited list of filtered criteria are highly time-consuming for users.

In the next section, based on the above-mentioned background, we are going to define the methodology and steps that we have followed throughout the study to carry out our research. That is to say, we will describe how to help creators of digital resources to provide better search services using visualization techniques, so that users can effectively find the resources they are looking for.

METODOLOGY

Stages and activities

Our research analyzes various approaches for visualizing information in a collection of digital resources, and aims to analyze the differences between them. The final aim is to draw conclusions that will enable the identification of advantages and disadvantages of each solution. To reach this goal it was first necessary to revise the state of the art (as we have just presented in the previous section) of information visualization, visualization techniques, digital repositories, and finally of metadata as a medium to improve search and access to digital resources.

Various types of techniques have been used to assess the relevance -from the design perspective- of different data visualization interfaces (Herman, Melancon, & Marshall, 2000).

In order to analyze and design search interfaces based on visualization techniques, firstly we browsed through digital resources related to a branch of knowledge of the Art and Architecture Thesaurus/AAT of the Europeana digital library (Gaona et al., n.p). To study the use of search interfaces and their relation to a collection of digital resources, we used the Europeana digital library as a case study. Europeana was chosen because of the following characteristics: i) it uses a semantic model for data exchange, the Europeana Data Model/EDM (Doerr et al., 2010), ii) it includes a large number of digital resources related to the European cultural heritage (over 40 million digital resources to date), iii) it uses the largest group of content providers on a European level, and finally iv) it allows reusability of open digital resources. Thus, it is an initiative that offers the opportunity to reuse digital resources, allowing teachers or professionals of the sector of cultural heritage to use them as learning objects for educational purposes. On the basis of these characteristics, we browsed through the Europeana digital resources to determine its level of coverage concerning a branch of knowledge of the Art and Architecture Thesaurus/AAT.

This study allowed us identify firstly, to what extent Europeana covers certain topics or areas of knowledge, permitting to reuse and develop learning objects of knowledge areas related to the European cultural heritage. In addition, we examined the relationships of part of the metadata elements of these digital resources, to verify the integrity of the search results of the users, and the availability of accessible digital resources. As this process revealed, the metadata of the majority of the Europeana digital resources we browsed, were insufficient. This enabled us to identify a number of deficiencies in the search processes of digital resources of a specific field of knowledge. On the basis of these results, we defined a set of activities related to the design of an alternative graphic search interface based on a knowledge representation scheme by using visualization techniques.

Part of the strategies defined in this research, concern a series of steps to be taken to develop search interfaces based on visualization techniques, evaluating them through a usability test, and finally providing recommendations to facilitate the development of visual search interfaces, using information visualization. Next, we describe in a simplified manner each of the stages of the present research. The stages I to IV can be considered associated to the Methodology

phase, while the rest are more related to the Results (VI stage) and Recommendations (VII stage) phases.

- I. Analysis and selection of a knowledge area
- II. Recognition of digital resources through browsing and extraction process
- III. Organization and classification of digital resources to analyse the coverage of a knowledge area (thematic coverage)
- IV. Search criteria based on metadata analysis
- V. Browsing and design of interfaces
- VI. Selection of the usability test and analysis of results
- VII. Recommendations for developing search interfaces through visualization techniques

I. Analysis and selection of a knowledge area

Firstly, we identified the area of knowledge that would be chosen in order to perform a coverage analysis of the Europeana digital resources. We adapted the term coverage (or thematic coverage), as it is defined by Codina (2000) and Whitehall (1992, 1995) to refer to the number of available digital resources related to a topic or knowledge area. Thus, coverage is analyzed to explore the completeness of the Europeana digital library in terms of topics or knowledge areas that include a larger number of digital resources. Therefore, to study the coverage of the library, we take as a basis a set of terms related to the topics of "styles and periods", specifically the "styles and periods by general era" and "styles and periods by region", i.e. knowledge branches of the AAT thesaurus. Next, we present the criteria according to which we selected the AAT thesaurus: i) the comprehensive and detailed representation of its hierarchical structure (Tudhope, Binding, Blocks, & Cunliffe, 2006), ii) the domain of knowledge (art and architecture) of the AAT thesaurus is one of the most complete and widely renowned (Aitchison, Gilchrist & Bawden, 2001), and finally, iii) the variety of conceptual descriptors it offers, which enables indexing and very large and precise queries (Soergel, 1995).

On the other hand, the specific domain of knowledge was selected according to the following factors: i) the topics of the field of "styles and periods" of the AAT thesaurus, are directly related to the domain of the Europeana digital resources concerning the cultural heritage ii) the terms of the AAT are related to a set of structured vocabularies concerning art and material culture (Weinberg, 1995), iii) the AAT thesaurus uses a "mature" knowledge representation scheme (Tudhope, et al., 2006). However like we discuss in the recommendation and conclusions sections, we can use a simple representation scheme in other knowledge area in order to establish a navigational structure of search.

Once we identified the domain of knowledge to be explored, it was crucial to know the Europeana data representation model and to identify the metadata elements used in each digital resource to proceed with the extraction process. Therefore, we used as reference the data exchange model defined by Europeana, known as the Europeana Data Model/EDM (Doerr, et al., 2010). This data exchange model has as its main objective to maintain original metadata as provided by the content providers, and to be a model with a wide semantic expression

capability. It is based on the best practices of the Semantic Web, and is compatible with standards such as the Open Archives Initiative - Object Reuse and Exchange protocol (OAI-ORE) (Lagoze et al., 2007) for Internet resource sharing and interoperability.

II. Recognition of digital resources through browsing and extraction process

In this second stage, we developed a strategy to link the Europeana digital resources to a specific knowledge area using selected terms of the AAT thesaurus. For this reason, we developed a Web Crawler to browse digital resources discovered in Europeana for each AAT thesaurus term. A Web Crawler analyzes the syntax of a Web page and extracts information according to its structure (metadata) (Tadapak, Suebchua, & Rungsawang, 2010). This extraction strategy was selected according to a series of features defined in a preliminary study we carried on relevant tools. These criteria were related to whether the tool: provides a clear definition of its architecture (Baeza & Castillo, 2004; Tripathy & Patra, 2008), is highly scalable (Boldi, Codenotti, Santini, & Vigna, 2004), favours optimizing processes (Edwards, McCurley, & Tomlin, 2001) and is highly efficient for content extraction processes (Castillo, 2005). On the other hand, metadata extracted from each digital resource is selected according to the elements of the Europeana Data Model/EDM.

III. Organization and classification of digital resources to analyse the coverage of a knowledge area (thematic coverage).

In the third stage we analysed the relationship between the extracted digital resources and the thematic areas selected for browsing. For this, with the support of a group of experts in cultural heritage, we reviewed the extracted results and found that most of the extracted resources did not correspond to the knowledge domain defined by the terms of the AAT thesaurus. For this reason, we designed a browsing strategy that used refined searches and keywords related to the domain of knowledge of each term of the AAT thesaurus. This strategy produced more relevant results as far as the knowledge area of each term of the thesaurus is concerned.

As a result, we found that a large set of topics associated to the AAT thesaurus knowledge area "Styles and Periods" was not fully covered by Europeana. On the other hand, the variety of results displayed while browsing suggested that Europeana did not support search methods to relate resources to a specific knowledge area.

To verify that, we had to perform various tests, analyzing the metadata quality according to evaluation criteria related to the completeness (identify the percentage of fields whose content was a non-zero value), and accuracy (identify the precision with which a digital resource was described). One possible explanation for this variety of results could be the poor quality of metadata definition in the digital resources. Therefore, these results were used to carry out a metadata quality analysis in order to: i) study the completeness and accuracy of the metadata extracted, and ii) identify search criteria associated with those metadata elements that present a higher degree of completeness, as we will soon explain in the next stage.

The results of this analysis showed a lack of completeness of the metadata defined according to the Europeana Data Model/EDM. However, such results are not new, since other studies concerning digital repositories such as ARIADNE (Ternier et al., 2009), the National Science Digital Library (Fox, Gonçalves, & Kipp, 2002) and other collections of digital resources (Bui, 2006), also found deficiencies concerning the definition of metadata. In fact, this deficiency, the lack of precision in the definition of metadata elements, is one of the main factors that directly influence digital resources search. Similarly, the absence of metadata elements for the classification of a digital resource according to topics or knowledge areas was reflected in the poor search results that we obtained through the study of thematic coverage. Those are factors that influence search results, and have also been mentioned by Cechinel (2009).

IV. Search criteria based on metadata analysis

The fourth stage was based on previous studies concerning various scenarios within the field of digital libraries (Ruecker, Radzikowska, & Sinclair, 2011). We also studied metadata, since it facilitates search processes for digital resources (Chirita, Gavriloaie, Ghita, Nejd, & Paiu, 2005; Hughes, 2005; Lytras & Sicilia, 2007; Moen, Stewart, & McClure, 1997; Ochoa & Duval, 2006; Park, 2005; Sanchez-Alonso & Sicilia, 2005; Stawniak, 2012). Therefore, to define the visual search criteria, we conducted a preliminary study on the metadata quality of the extracted digital resources related to their completeness. The concept of completeness, as we have just said, relates to the number of the fields whose content is a non-zero value (Moen, et al. 1997; Ochoa & Duval, 2006; Park, 2005). In the case of multi-valued fields, a field is considered complete if there is at least one recorded value. According to authors such as Ochoa & Duval (2006) this metric can be used as a first indication for the evaluation of the completeness of a metadata record. Following this process, the metadata that presented a high percentage of completeness were directly related to metadata elements defined by Europeana. These sets of metadata are: provider, country, language and type of digital resource. Having these results as a basis, this set of metadata was used as a reference to define the criteria used for the visual search strategies that were implemented on *knowledge representation schemes*.

V. Exploration and design of interfaces

At this stage navigation interfaces were implemented as strategies of visual search. To do so, we relied on several design criteria associated with the development of visual interfaces and with the use of knowledge representation schemes based on taxonomies and thesauri. As far as the design of the interfaces is concerned, principles of HCI were taken into account to facilitate navigation and interaction with the user. We identified the following tasks:

- *Browsing*: Identifying several browsing methods, which according to Marchionini (1997) allow the definition of strategies for information finding, and according to Lin (2010) allow the visualization of large amounts of information and the perception of structures and hierarchical relationships. According to Marchionini (2006), these browsing methods, when they are used in a semantically rich interface, can help users to formulate questions and to browse having a comprehensive view of the available digital resources.

- *Navigation*: Displaying information through maps of categories based on the taxonomic structure. This activity facilitates the navigation of contents by domain of knowledge (Card, et al., 1999) to provide users the location of certain terms.
- *Interaction*: Zooming in or out on the navigation map, and panning to run through the whole structure of the terms that could be consulted. Through the zoom strategy, users apply focus and context techniques (Lamping, Rao, & Pirolli, 1995) which allow focusing on the node that is currently displayed, as well as unfocusing nodes related to other levels of hierarchy.

Given these criteria, we started using search strategies based on browsing. These strategies were also supported by navigation activities based on graphic visualization techniques, with the aim to take full advantage of the users browsing capabilities (Marchionini, 2006). Browsing strategies have been previously studied under several work scenarios giving good results (White, Kules, & Drucker, 2006; White & Roth, 2009). White (2006) highlights the importance of these search strategies, since they enable users to query and navigate within a defined structure, thus promoting learning and research. These results are supported by studies related to the browsing and navigation level of taxonomic structures (Uddin & Janecek, 2007; Wang, et al., 2008), as well as by the importance of search tasks based on learning, as the most appropriate strategies to perform navigation based activities (Marchionini, 2006).

One factor for the design of interfaces were the principles that determine the use of taxonomies, to discover ways to improve our taxonomic representation. These principles were related to: i) the design of a simple representation scheme and fine-grained categories (Morante, 2003) to provide a usable navigation structure, ii) the levels of navigation, to analyze the complexity of a structure and navigation path (Bruno & Richmond, 2003), and iii) the browsing levels in order to identify the best navigation structure according to the depth of the taxonomy (Wang, et al., 2008). Based on these principles, in addition to the design of visual interfaces we took into account two key concepts. The first is the idea of visually rich interfaces, where the representations of each element of a collection of digital resources are complemented with emerging search tools (Ruecker, 2003). The second is to design search interfaces based on the use of thesauri (Shiri, Revie, & Chowdhury, 2002), and more importantly the following:

- Integrating the thesaurus and document retrieval on screen
- Using flexible forms for the selection of query terms
- Providing hierarchical and alphabetical lists to support different strategies
- Describing a term and its location within a hierarchical structure

For the definition of the taxonomic structure we used well-defined thesauri. This is very useful for our research since they provide: i) a clear navigational structure, ii) a well-defined hierarchy, and finally iii) wide semantic capabilities.

Finally, given the purpose of our study and taking into account the above-mentioned principles followed for the development of visual interfaces, we selected visualization techniques

according to: i) graphical representation of structures (Herman, et al., 2000), ii) thematic clustering strategies (Tanja Merčun, Žumer, & Aalberg, 2012) and, finally iii) hierarchical representation of categories (Schulz, Hadlak, & Schumann, 2011; Shneiderman et al. 2000). Based on these selection criteria, we developed eight interfaces based on eight different information visualization techniques, visually classified into four types of strategies: *radial* (*radial*, *radial-search*, *relation*), *hyperbolic* (*hypertree*, *sunburst*), *tree* (*tree*), and by category (*icicle*, *treemaps*) (see Figure 1). We later focused our usability study on evaluating the visual perception capabilities of each of these 8 interfaces at various levels: attention, retention and understanding of topics associated with an area of knowledge, and retention associated with the number of resources and their thematic coverage. Furthermore, interaction was evaluated at navigation and classification levels using taxonomic structures that were implemented through a set of topic-related terms of "styles and periods" as they are defined in the AAT thesaurus. Finally, we developed a prototype of a visual search interface, which combines visualization techniques with a framework that allows linking text searches and displaying digital resources using metadata criteria. These visual interfaces were selected according to a usability study conducted by a group of 32 persons.

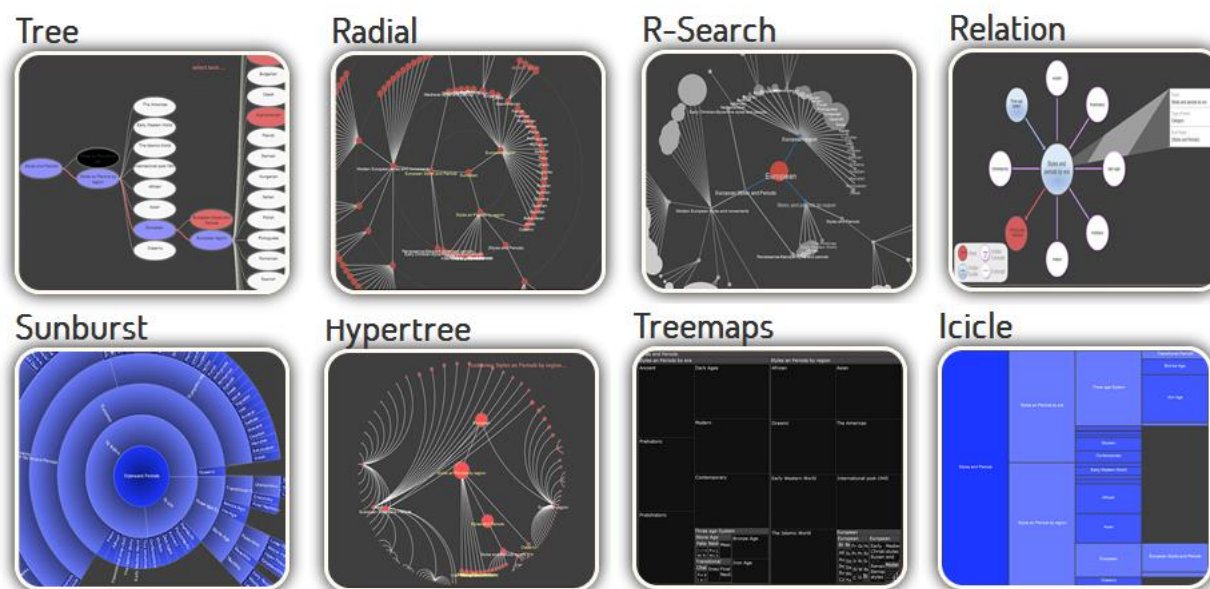


Figure 1. Implementation of search interfaces

RESULTS

VI. Selection of tests and results

The sixth stage of this research included the experimental design of visual interfaces, as well as a series of usability tests run on a predetermined knowledge area or topic. These tests included various stages towards the analysis of activities associated with: i) visual perception, and ii) interaction that will be analysed in detail in the forthcoming sections.

a. Visual perception test

For the visual perception tests we consulted usability experts who collaborated for their design. Before conducting the tests, the moderator (a usability expert) explained briefly the test process, indicating its purpose and the type of activities that each participant had to perform. The moderator first presented to each participant the work scenario, so that the participant did not have pressure during the activity.

To perform the tests, we used an adequate workplace for participants and basic recommendations for adapting the space as proposed in Bevan (1994) (e.g. desk, video recorder, laptop, soundproof rooms). The term visual perception, in our first usability test, refers to the ability of a user to interpret information encoded and represented through a visual image (Cleveland, Diaconis, & McGill, 1982). In such studies, the impact assessment allows designers to optimize visualization to improve their design (Mackinlay, 1991; Mackinlay, Hanrahan, & Stolte, 2007).

Taking into account relevant studies (Cleveland, et al., 1982, Mackinlay, et al., 2007), and having the support of usability experts, the design of visual perception test aimed to provide participants a simulation of the visualization techniques of the search interfaces, based on the representation of a set of images that alluded to each technique. Thus, each image (scenario) represented a hierarchical navigation structure of a visualization technique. To achieve that, these scenarios were designed taking into account factors commented in previous studies (Lidwell, Holden, & Butler, 2010; J. Nielsen, 1994b) and were based on the following criteria established by usability experts:

- *Simplicity of topics*: Each scenario was designed to represent simple examples using topics associated to areas of knowledge such as geography, sports, movies, drinks, animals, fruits, home, and meals (one image for each visualization technique).
- *Hierarchical structure*: The hierarchical structure of all examples should be the same (3 levels deep).
- *Number of terms*: Each scenario should display the same number of terms (12 to 13 terms).
- *Thematic association*: Each term of the scenario or taxonomic representation, should be related to the whole scenario and to each level of the established hierarchy.

One of the most frequently used methods to carry out visual perception tests is the so called "five-second test" (Boren & Ramey, 2000; Dumas, 1999; J. Nielsen, 1994b). According to this method we showed to the participant a graphical user interface/GUI during five seconds, after which an expert makes a series of questions to the participants in order to get their first impressions about the design. In our study we used this method for the assessment of four aspects (attention, retention of information, understanding and subjective impression):

- *Attention*: If the user was able or not to identify the topic that the image represented.
- *Retention of information*: The number of terms that the user could remember.
- *Understanding*: Terms were connected on three levels. The moderator tried to identify in what way the depth levels could be recognized by each user.
- *Subjective impression*: The participant was requested to rank all the interfaces after having experienced them.

Another goal of this study was to familiarize the participants, through the visual association of the taxonomic structure, with the hierarchical scheme of each interface before interacting with the interfaces and carrying out search processes. At the same time, this test focused on the evaluation of each activity by a moderator, a skilled person who was responsible for developing the test. The moderator made a number of questions related to each image in order to assess the level of perception and the subjective impressions of users, asking them to indicate, for example, the terms of the picture they could better remind, or making questions such as “what relations can you remember?” (this was done for each depth level presented in the image) or “which was the general topic of the picture?”

b. Interaction test

For the interaction test, we carried out a case study concerning interface interaction. Users, that played as if they were students, should decide whether the application helped them to locate a number of terms and whether the results of the search were sufficiently covered for each indicated period, judging by the number of digital resources displayed during the query process. Thus, each participant had to find a set of terms located on different levels of hierarchy within an interface, and perform a series of activities in order to: i) identify a term within the taxonomic structure, ii) identify the number of resources associated with each term, iii) determine up to what point the term was covered, according to the number of digital resources found (thematic coverage), and iv) assess the interface according to usability criteria related to ease of use, navigation, classification and aesthetics.

c. Analysis of the results

An analysis of the final results was carried out after the completion of the above-mentioned visual perception and interaction tests. When comparing the subjective evaluation results (visual perception test) with the results of the activities that each participant had to perform (interaction test), in some cases participants had difficulty in identifying the number of digital resources related to the terms. Participants expressed their frustration when they could not find the terms that the moderator asked.

Taking into account the relevance of the results obtained in terms of perception and interaction, a complementary study was carried out in order to verify the results about the thematic coverage, to analyze whether this criterion visually affected access to a collection of digital

resources. Preliminary results showed that interfaces presenting different sizes for each node, caught the attention of the participants allowing them to identify which terms presented a bigger or smaller number of resources. These visual effects allowed participants to distinguish topics with a big coverage to those with a smaller within a hierarchical structure, fact that facilitated access to a digital resources collection.

The results of the study of perceptual capabilities, showed that the level of retention is not a condition that affects directly the recognition of a thematic area within a taxonomic structure. Similarly, retention does not directly affect the reasoning of a user when it comes to identifying relations at different levels of the taxonomic structure, i.e. thematic relations linked on different levels of a taxonomy. Moreover, the level of satisfaction concerning the subjective impression (visual perception) of the participants for each of the interfaces, proved to be an important factor related to the role of the visualization of thematic coverage within a taxonomic hierarchy.

Conducting the perception tests prior to the use of the interfaces, we allowed participants to get familiar with the interfaces they later had to use for searching and browsing. However, in some cases the evaluation according to the visual perception test differed from the results of the final interaction test. It is important to note that the quality of the taxonomic structure could affect the use of visual interfaces and thus, so we preferred to use simple taxonomies for these tests.

The participants generally considered the visualization techniques of the study useful for enhancing access to a collection of digital resources through a taxonomic hierarchy. However, according to opinions expressed by some participants during think aloud protocols, the visual navigation had to be complemented by textual search.

In summary, users prefer methods that allow them to find rapidly a concrete resource within a taxonomic structure, by using search criteria based on specific topics/knowledge area (thematic coverage). However, the learning curve (related to the time spent to understand the use of the interfaces) is a very important variable for the proper use of such strategies. In order to maintain this learning curve, it is important to define, on an aesthetic level, visual alternatives that can provide a better understanding of classification, location and access based on knowledge representation schemes. Better response time and query performance for a large volume of digital resources are elements that could also favour the usability of an interface and that should be considered.

RECOMMENDATIONS

VII. Set of recommendations for developing search interfaces using visualization techniques

In order to improve the development of search interfaces by using information visualization techniques, we have defined a set of recommendations for repositories.

- *Metadata Quality Management*: Defining policies for metadata quality management and administration, to improve the precision of metadata definitions, and providing search results that include relevant digital resources.
- *Linking metadata fields associated by topics and concepts*: Defining strategies to facilitate search of digital resources by topic or knowledge area. This could be accomplished by using LOM metadata definitions, such as "classification", "relation", or the use of keywords in order to link concepts as related terms, associated concepts, preferred term, non-preferred term, etc. Such metadata would offer classification by topics or knowledge areas to simplify the access to a digital resources collection.
- *Knowledge representation scheme*: Selecting a knowledge representation scheme comprising at least:
 - *Taxonomic Hierarchy*: A basic definition of its taxonomic hierarchy, to facilitate the classification and use of visualization techniques.
 - *Relations of hierarchy and association*: Establishing an association or hierarchy relation within the taxonomic structure to link related terms (generic or specific) during the search process.
 - *Enriched vocabulary Management*: Providing alternative expressions based on complementary terms that include synonyms, non-preferred terms, etc.
- *Defining search criteria*: Depending on the strategies of the repository, defining search criteria based on metadata related to topics or knowledge areas to facilitate browsing and user interaction, and integrating criteria related to the educational context, user profiles, language and types of digital resources, etc.
- *Usability Testing*: Defining and establishing usability tests prior to the development of a solution, to measure both user perception and interaction, with the aim to improve the quality of the design and the functionality of the interfaces, and thus improving search processes and access to relevant digital resources. Similarly, it is recommended regularly to repeat this test once the interfaces are operational. This process significantly enriches the use and adequacy of complementary strategies that improve access to digital resources of a repository.

Although not all visualization techniques facilitate the process of finding digital resources, they offer a number of features that improve accessibility and user satisfaction. Therefore, the selection of appropriate techniques goes beyond the functionality they could offer, since most of them aim to provide alternative browsing by the use of taxonomies. However, it is necessary to identify a number of technical factors that affect the developer tools and function libraries in order to facilitate their implementation in interfaces. To this direction, we propose the following recommendations:

- *Selecting graphic function libraries*: Knowing the graphic computing capabilities of the library in order to determine the visualization techniques it supports, and identify the graphic function libraries that allow the integration of add-on interactions that facilitate their use.
- *Response capacity*: Measuring the response time and the performance of a visualization technique and of a given library, according to the knowledge representation scheme of the repository. This can measure the navigation response time for loading categories and nodes associated to each level of depth, depending on the complexity of the taxonomy used.
- *Format integration and support*: Selecting function libraries which support JSON, XML, RDF data formats for the development and implementation of the taxonomy, according to the knowledge representation scheme defined in the repository. Currently, most function libraries support taxonomic structures in JSON format; however these taxonomic structures can also be integrated with digital resources in order to allow querying XML, RDF or SKOS formats, by linking digital resources as Linked Data (Bizer, Heath, & Berners-Lee, 2009).
- *Scalability and work environment*: Offered by the tools in order to adapt to changes and updates that the knowledge representation scheme can undergo because of new digital resources are added to the repository.
- *Navigation Paths*: Defining adequate navigation paths to guide browsing within the navigation structure and the location of digital resources related to specific topics.
- *Integration of supplementary searches*: Integration of complementary search strategies based on access features defined by educational context (secondary education, advanced, continuing, university, professional), user profiles (teachers, students, professionals), digital resource format (image, video, RTF document, presentation, slides, etc.) or digital resource language.
- *Display of digital resources*: Adding display strategies for relevant digital resources using a paging mechanism, based on digital resources evaluation by users, taking into account their profile as assessed by experts, registered users, professionals, students, etc.

CONCLUSIONS

It is important to remark the use of information visualization techniques in educational repositories as an alternative to improve “traditional search”. These techniques provide a number of tasks, as follows:

- i) *Visual perception*: Improve the search and access to a set of educational resources by regular users (students and/or teachers). Visual perception facilitates understanding of

information associated with the attention capacity in order to identify the classifications associated color, size, relationship and shape of the navigation structure.

- ii) *Navigational search*: To provide the user a clearer idea of topics and knowledge area of search in its entire full context. This facilitates a better interpretation and assimilation of the entire knowledge area.
- iii) *Training and learning*: Better understanding of topics or knowledge areas related to the search. In particular this last issue has been mentioned by Davis & Connolly (2007) and Swan & Carr (2008) like a challenge of current educational repositories, generating a progressive abandonment of search process.
- iv) *Semantic search*: Improve the search of learning objects through the use of knowledge representation schemes like ontologies and thesauri. Knowledge schemas facilitate the understanding of information through the deployment of different navigation hierarchies, classification and thematic coverage within a taxonomic structure.
- v) *Use of metadata*: Promote the use of metadata in repositories to carry out the access of relevant learning objects according to strategy of search such as language, formats, content provider or copyright.
- vi) *Use of linked data*: Improve the access and use of learning objects making use of a number of benefits derived from the use of linked data, e.g., through the use of associate relations and hierarchy defined in knowledge representation schemes.

in particular each selected visualization technique, facilitates the integration of taxonomic structures based on the use of graphics libraries. In this direction, we should define as future lines of work, several issues of visualization techniques in order to allow creators of digital repositories, the integration of visual search interfaces from several point of view:

- i) *Navigational structure*: Define a simple hierarchical structure, or define different levels of complexity of use according to user profiles.
- ii) *Forms to represent data*: Define alternatives of navigational search according to its structure (nodes, squares, circles). The visual integration of the number of digital resources available to each topics (thematic coverage) is a factor that significantly would improve the access to a collection of digital resources, as long as the results are esthetically pleasing appearance to the participants. This condition is one of usability principles, which must contain all search interfaces at visual perception level.
- iii) *Hierarchical relationships*: Define levels of relationship and association between digital resources and concepts of knowledge in taxonomic structure. It is important to remark the use of simple navigational path to identify the hierarchical level of all navigational structure.

- iv) Ease of use: For novice participants of repositories it is good to use a simple and familiar navigational structure like the tree interface. It is also important to remark the establishment of usability test in order to add more components to improve the search (e.g. navigational path, textual search, number of digital resources, keywords, related terms etc.). However, depending on hierarchy levels or depth that repository provided, try to work with user-friendly interfaces: (tree, *radial* or *icicle*) that support the use of navigational paths in order to identify the classification level within the taxonomic structure.

RELATED RESEARCH PUBLICATIONS

Next we present the results of this research in publications.

JOURNALS

- 1) Gaona, P., Martin, D., Feroso, A., Sánchez, S. (2014) A usability study of taxonomy-visualization user interfaces in digital repositories. Online Information review. Volume 38, Issue 2. Impact Factor: 0,932

Summary: This paper reports on a research on the use of different types of search and navigation interfaces in digital repositories. Our aim was to analyse user interfaces to support search and collection visualization and navigation from a usability perspective, and to identify which are more successful. As the final objective, we wanted to offer repository owners enough scientific ground to support their decisions when they face the task of choosing an interface that can really help users to effectively locate and visualize resources over large digital collections.

- 2) Gaona, P. Feroso, A., Sánchez, S. (in progress). Exploring the coverage of the Europeana digital library using the AAT Thesaurus as a use case.

Summary: The purpose of this paper is to analyze the completeness of the Europeana digital library in terms of coverage by topics, using a concrete knowledge area of the Art & Architecture Thesaurus® - AAT as an example use case. The paper studies the relevance of the results provided by queries targeting the resources' metadata elements defined by the EDM Europeana model, and introduces a quality accuracy analysis of some relevant metadata elements in a small sampling of digital resources of Europeana. The main aim is to assess if Europeana can be really seen as the new Alexandria Digital library from the perspective of teachers, academics and researchers that might want to retrieve and reuse those resources for learning purposes.

RESULTS OF RESEARCH STAY IN ATHENS

Next is part of the results of research stay in Athens conducted between November 2013 and February 2014.

- 1) Gaona, P., Stoitsis, G. Sánchez, S., Biniari, K. (in progress). Visual framework for the access of several repositories through a knowledge representation scheme: Study case of Organic.Edunet and VOA3R/AGRIS.

Summary: The main goal of this paper is to offer an alternative method of visual search with the aim of facilitate access of academic learning objects for several repositories through the use of information visualization. To carry out this purpose, we development a visual framework tool and took as a study case Organic.Edunet and ABN/AGRIS repositories to improve the access of academic and scientific publications resources respectively.

PROCEEDINGS

- 1) Martin, D., Sánchez, S., Gaona, P. & Marianos, N. (2013). Applying visualization techniques to develop interfaces for educational repositories: the case of Organic.Lingua and VOA3R. *Proceedings of the Learning Innovations and Quality: The Future of Digital Resources*, 60-67.

Summary: This paper explores the fundamental reasons for problems related to usability at the light of the results of several usability studies carried out in the context of the European projects “Organic.Lingua”(www.organic-lingua.eu) and “VOA3R” (www.voa3r.eu), where two different digital repositories are begin developed.

- 2) Gaona, P., Sánchez, S., Feroso, A. (2012). Análisis de cobertura del tesoro AAT en la biblioteca digital Europea: ideas preliminares para su uso en Educación. Multidisciplinary symposium on the design and evaluation of digital content for education. SPDECE

Summary: This paper present the preliminary results of thematic coverage of Europea digital library according to the topics and knowledge area of “styles and periods” present in AAT thesaurus.

PAPERS IN NATIONAL PUBLICATIONS

- 1) Gaona, P., Sánchez, S. (2014) Information Visualization: A proposal to improve search and access digital resources in repositories. *Ingeniería e investigación*. ISSN: 2248-8723. Volume 34, Issue 1. Impact Factor: 0,034

Summary: This paper focus in proposal best practice to search and access relevant digital resources in repositories through the use of visualization techniques. The paper

presents a tool prototype as one possible approach to represent and search digital resources in learning object repositories.

- 2) Gaona, P., Sánchez, S., Feroso, A. (in press) Visualization techniques through search interfaces in learning object repositories. TECNURA. ISSN: 0123-921X. Volume 18, Issue 41.

Summary: The paper presents preliminary results in order to evaluate the most representative aspects to carry out the development of visual search interfaces of learning objects in digital repositories, according to principles of visualization techniques and hierarchic taxonomic classification.

- 3) Gaona, P., Sánchez, S. (2013). Modelo Informático para extracción de recursos digitales en Internet. TECNURA. ISSN: 0123-921X. Volume 17, Special Issue.

Summary: This paper presents the result of a tool in order to extract digital resources of Europeana digital library according to knowledge area of styles and periods by the AAT thesaurus.

Keynotes:

- [1] Available online at: <http://www.merlot.org/>, (Accessed 7 Mayo 2014)
- [2] Available online at: <http://portal.mace-project.eu/>, (Accessed 7 Mayo 2014)
- [3] Available online at: <http://www.ariadne-eu.org/>, (Accessed 7 Mayo 2014)
- [4] Available online at: <http://209.87.56.111/drupal/>, (Accessed 7 Mayo 2014)
- [5] Available online at: <http://eprints.rclis.org/>, (Accessed 7 Mayo 2014)
- [6] Available online at: <http://www.organic-lingua.eu/es>, (Accessed 7 Mayo 2014)

Índice General

Agradecimientos	I
Resumen	III
Abstract.....	IV
Executive Summary	V
Índice General.....	XXI
Índice de Figuras	XXV
Índice de Tablas.....	XXVII
 Capítulo 1: Introducción	 2
1.1 Panorama de trabajo	2
1.1.1 Objetos de aprendizaje	2
1.1.2 Metadatos de objetos de aprendizaje.....	3
1.1.3 Repositorios digitales educativos	4
1.1.4 Usabilidad	5
1.2 Motivación	6
1.3 Objetivos y contribuciones.....	8
1.3.1 Contribuciones	9
1.4 Método de trabajo.....	10
1.4.1 Estudio del estado de la cuestión	10
1.4.2 Identificación y descripción del problema	10
1.4.3 Definición del modelo	11
1.4.4 Análisis de calidad de metadatos	11
1.4.5 Definición de un framework de visualización	11
1.4.6 Estudio de usabilidad	12
1.4.7 Análisis de resultados	12
1.4.8 Formulación de conclusiones.....	12
1.5 Estructura del documento.....	12
 Capítulo 2: Estado de la Cuestión	 16
2.1. Esquemas de representación de conocimiento.....	16
2.2. Formas de representar el conocimiento en repositorios digitales.....	19
2.2.1 Taxonomías	19
2.2.2 Tesoros.....	22
2.2.3 Ontologías.....	23
2.3 Estrategias de acceso sobre recursos digitales	27
2.3.1 Métodos de búsqueda.....	28
2.3.2 Métodos de acceso	29

2.3.3 Criterios de búsqueda	31
2.3.4 Despliegue de resultados de búsqueda	32
2.4 Visualización de información en repositorios y bibliotecas digitales	36
2.4.1 Visualización de Información.....	36
2.4.2 Interacción persona ordenador.....	37
2.4.3 Principios de usabilidad.....	37
2.4.4 Visualización de información en bibliotecas digitales	39
2.5 Técnicas de visualización	40
2.5.1 Visualización mediante árboles	40
2.5.2 Visualización de tipo radial.....	42
2.5.4 Visualización por categorías	44
2.6 Usabilidad de la visualización y navegación de colecciones digitales	46
2.6.2. Usabilidad de las taxonomías y esquemas de representación de conocimiento	47
2.6.3 Revisión de estudios sobre usabilidad de interfaces de búsqueda	48
2.7 Resumen	50
Capítulo 3: Planteamiento del Problema	54
3.1 Deficiencias de los métodos de búsqueda y acceso sobre objetos de aprendizaje	54
3.2 Limitaciones de los esquemas de representación de conocimiento en repositorios.....	57
3.3 Limitaciones en el uso de metadatos	58
3.4 Resumen	60
Capítulo 4: Diseño de una solución.....	62
4.1 Modelo de trabajo	62
4.2 Análisis de datos: Caso de estudio biblioteca digital Europeana.....	63
4.2.1 Análisis y selección del área de conocimiento a explorar.....	64
4.2.2 Proceso de exploración y extracción de recursos digitales.....	66
4.2.3 Análisis de cobertura de recursos digitales explorados	69
4.2.4 Definición de criterios de búsqueda.....	71
4.3 Diseño de interfaces de navegación	73
4.3.1 Diseño de estructura jerárquica y proceso de transformación de datos.....	74
4.3.2 Selección de bibliotecas de funciones e implementación de técnicas de visualización sobre interfaces de navegación	76
4.3.3 Diseño de un framework basado en interfaces de búsqueda	84
4.4 Resumen	88
Capítulo 5: Metodología y diseño experimental.....	92
5.1 Método de prueba y evaluación de interfaces de búsqueda navegacionales	92

5.2 Selección de participantes y pruebas	92
5.3 Test de percepción	94
5.4 Test de interacción	98
5.4.1 Fase de entrenamiento	98
5.4.2 Fase de prueba	100
5.5 Resumen	104
Capítulo 6: Discusión y análisis de resultados.....	106
6.1 Análisis de los resultados del estudio de usabilidad	106
6.1.1 Percepción e impresiones subjetivas	106
6.1.2 Resultados de interacción.....	110
6.1.4 Resultados de precisión	112
6.1.5 Resultados sobre percepción de la cobertura temática	113
6.1.6. Comprobación de hipótesis.....	114
6.2 Análisis de calidad de los metadatos	119
6.2.1 Completitud.....	119
6.2.2 Precisión	122
6.3 Discusiones	124
6.3.1 Discusiones de resultados de usabilidad.....	124
6.3.2 Discusiones de resultados sobre análisis de calidad de metadatos	126
6.4. Elaboración de una propuesta	127
6.4.1 Análisis y verificación de esquema de representación de conocimiento.....	128
6.4.2 Selección de técnicas de visualización.....	130
6.4.3 Selección de herramientas y bibliotecas de funciones gráficas	132
6.4.4 Definición de framework de integración de búsquedas	134
6.4.5 Definición de despliegue de resultados y complementos de consultas.....	136
6.5 Desafíos y retos del área de visualización de información	136
6.6 Resumen	139
Capítulo 7: Conclusiones y trabajo futuro.....	142
7.1 Resumen de objetivos y resultados	142
7.2 Conclusiones.....	143
7.2.1 Asociados al área de visualización de información.....	143
7.2.2 Asociados a las técnicas de visualización	144
7.2.3 Asociados a los esquemas de representación de conocimiento	145
7.2.4 Asociados a los estudios de usabilidad	145
7.2.5 Asociados a los estudios de calidad de metadatos	146
7.3 Recomendaciones	147

7.3.1 Asociados a repositorios digitales	147
7.3.2 Asociados al diseño de las interfaces.....	148
7.4 Aportaciones originales	149
7.5 Trabajos futuros.....	150
7.6 Publicaciones derivadas de la investigación	151
7.6.1 Revistas internacionales	152
7.6.2 Congresos.....	152
7.6.3 Revistas nacionales indexadas	152
Anexo A	153
Estrategia de extracción de recursos digitales (normal vs refinada)	153
Anexo B	156
Cuestionario de datos prueba de usabilidad	156
Referencias Bibliográficas.....	159

Índice de Figuras

Figura 1. Tipos de conocimiento humano y su representación (Durkin, 1994).....	17
Figura 2. Descripción de un recurso digital metadatos - Ejemplo repositorio MACE.....	21
Figura 3. Estructuras y estrategias para organización del conocimiento (Lei, 2008)	25
Figura 4. Métodos de acceso sobre recursos digitales.....	28
Figura 5. Ejemplo de métodos de búsqueda textual repositorio VOA3R	30
Figura 6. Ejemplo Métodos de búsqueda navegacional repositorio Organic.Edunet	31
Figura 7. Ejemplo de criterios de búsqueda definidos en biblioteca digital Europea	32
Figura 8. Ejemplo del despliegue de recursos digitales tipo lista - textual MERLOT.....	33
Figura 9. Ejemplo despliegue de recursos digitales tipo lista, pre-visualización MACE	34
Figura 10. Formas básicas de representación de árboles (Graham & Kennedy, 2010)	41
Figura 11. Adaptación de técnica de visualización tipo árbol.....	41
Figura 12. Adaptación de técnica de visualización radial	42
Figura 13. Visualización sunburst.....	43
Figura 14. Visualización hypertree	44
Figura 15. Tipos de representación por categoría. Ejemplo de una faceta (a). Adaptación de técnica de visualización treemaps (b).	45
Figura 16. Adaptación de técnica de visualización icicle	46
Figura 17. Modelo de trabajo planteado.....	62
Figura 18. Metadatos presentes en el modelo de datos de Europea EDM	65
Figura 19. Modelo de extracción de datos definido para el Web Crawler	66
Figura 20. Modelo para la extracción de datos definidos en el Web Crawler.....	67
Figura 21. Metadatos extraídos de un recurso digital en Europea	68
Figura 22. Número de recursos digitales asociados a ciertas temáticas del tesauro	69
Figura 23. Análisis de completitud para conjunto de metadatos obligatorios y recomendados	72
Figura 24. Niveles de profundidad de la estructura jerárquica definida por el tesauro AAT	75
Figura 25. Proceso de análisis y transformación de datos.....	76
Figura 26. Interfaz de visualización tipo árbol.....	77
Figura 27. Interfaz de visualización “radial”	78
Figura 28. Interfaz de navegación “radial-search”	79
Figura 29. Interfaz “relation” (a) Nivel jerarquía superior (b) Nivel de jerarquía inferior.....	80
Figura 30. Interfaz de visualización “sunburst”	80
Figura 31. Interfaz de visualización “hypertree”	81
Figura 32. Interfaz de visualización “treemaps”	82
Figura 33. Interfaz de visualización “icicle”.....	83
Figura 34. Arquitectura del modelo de tres capas para el área de visualización.....	84
Figura 35. Integración de búsqueda textual y navegacional	85
Figura 36. Criterios de búsqueda visual mediante diagramas de tipo pastel	86
Figura 37. Despliegue de listado de recursos digitales.....	87
Figura 38. Descripción del recurso digital.....	88
Figura 39. Ejemplo de escenario “películas” a partir de interfaz treemaps	95

Figura 40. Interfaz de trabajo para realizar prueba de usabilidad de percepción visual.....	96
Figura 41. Momento de interacción de un usuario en prueba de usabilidad	103
Figura 42. Clasificación de las interfaces de acuerdo a la percepción de los participantes	106
Figura 43. Ejemplo de niveles de relación para interfaz radial-search.....	108
Figura 44. Nivel de comprensión por interfaces	109
Figura 45. Niveles de profundidad para localizar el término “art deco” interfaz radial-search .	110
Figura 46. Porcentaje de éxito y fracaso para la localización de términos en interfaces.....	113
Figura 47. Regresión lineal entre cobertura temática y precisión	118
Figura 48. Etapas y consideraciones para el desarrollo de interfaces visuales a partir de esquemas de representación de conocimiento	128
Figura 49. Criterios de evaluación para selección de esquema de representación de conocimiento.....	129
Figura 50. Criterios de evaluación para selección de técnicas de visualización	130
Figura 51. Resultados de evaluación para selección de técnicas de visualización	131
Figura 52. Criterios de valoración herramientas de visualización	134
Figura 53. Arquitectura del modelo propuesto de tres capas para el área de visualización	135

Índice de Tablas

Tabla 1. Diferencias y funcionalidades de esquemas de representación de conocimiento	26
Tabla 2. Funcionalidades de instrumentos de representación de conocimiento	27
Tabla 3. Estrategias usadas por repositorios digitales para acceder a recursos digitales.....	34
Tabla 4. Nivel de cobertura temática términos del tesoro AAT explorados en Europeana	70
Tabla 5. Listado de metadatos con mayor porcentaje de completitud	73
Tabla 6. Perfil de participantes.....	93
Tabla 7. Especificaciones de equipo y software de captura	93
Tabla 8. Instrumento para registrar interacción de percepción de usuarios	97
Tabla 9. Instrumento para registrar fase de entrenamiento de interacción de usuarios.....	99
Tabla 10. Instrumento para registrar fase de prueba de interacción de usuarios.....	100
Tabla 11. Resultados estadísticos asociados a la capacidad de percepción.....	107
Tabla 12. Estadísticas sobre impresiones del usuario sobre aspectos de usabilidad	111
Tabla 13. Promedio estadístico de valoración de interfaces a partir de cobertura temática	114
Tabla 14. Interfaces de prueba de homogeneidad utilizando HSD - técnica de Tukey	115
Tabla 15. Prueba distribución normal de Kolmogorov-Smirnov sobre todas las muestra	116
Tabla 16. Análisis estadístico de atributos mediante correlación de Pearson (r)	116
Tabla 17. Coeficiente de regresión lineal para cobertura temática	117
Tabla 18. Resultados de completitud sobre elementos obligatorios	120
Tabla 19. Resultados de completitud sobre elementos recomendados.....	121
Tabla 20. Resultados de completitud sobre elementos opcionales	121
Tabla 21. Resultados de precisión sobre recursos digitales seleccionados (5.500).....	123
Tabla 22. Características de herramientas de visualización open Source	132
Tabla 23. Resumen de objetivos y resultados.....	142
Tabla 23. Estrategia de extracción recursos digitales (Normal VS Refinada)	153

Capítulo 1

Introducción

El propósito fundamental de este capítulo es ofrecer una síntesis de los temas de trabajo que abordaremos a lo largo de esta memoria. A continuación nos centraremos en el contexto y la motivación que ha dado lugar al desarrollo y formulación del presente trabajo. Posteriormente, se exponen los objetivos que se pretenden abordar, describiendo las contribuciones más representativas, seguido del método de trabajo utilizado. Finalmente se resume la estructura que se irá desglosando a lo largo del documento.

Capítulo 1: Introducción

*“Desde la idea a la palabra; desde la palabra a la idea”
(Casares, 1942)*

1.1 Panorama de trabajo

La presente investigación involucra varias áreas de estudio enfocadas al análisis y visualización de objetos de aprendizaje a partir de interfaces de búsqueda sobre repositorios digitales. En este apartado se lleva a cabo de manera breve la relación de las áreas más representativas, con el propósito de ofrecer un panorama de las temáticas que se abordarán lo largo del documento.

1.1.1 Objetos de aprendizaje

De acuerdo a la IEEE, (2002) *“un objeto de aprendizaje se puede definir como una entidad, digital o no, que puede ser usada para el aprendizaje, la educación o la enseñanza”*. Esta definición ha sido uno de los puntos de referencia, que ha permitido abordar el concepto desde varios enfoques en el ámbito académico. Algunos de ellos mediante el uso de atributos relacionados con: i) aspectos didácticos (Polsani, 2006), ii) entornos de aprendizaje (Longmire, 2000; Wiley, 2002a), iii) propiedades de accesibilidad e interoperabilidad (N. Friesen, 2001); y iv) perdurabilidad (Rehak & Mason, 2003). Otras definiciones surgieron a partir del uso de metáforas, algunas de las más representativas asociadas con el uso de átomos (Wiley, 2002b) y bloques de construcciones en Lego (Hodgins & Conner, 2000). No obstante, la mayoría de autores coinciden en el desarrollo de objetos de aprendizaje con el fin de ofrecer características modulares, modificables y fáciles de reutilizar (Duncan, 2009; Polsani, 2006; Sanz, Dodero, & Sánchez, 2011).

El concepto de objeto de aprendizaje ha permitido, entre otras cosas, generar prácticas basadas en compartir materiales educativos abiertos (Schaffert & Geser, 2008). Una de las motivaciones que dio lugar a este tipo de prácticas, fue el desarrollo de recursos educativos abiertos de calidad, con el propósito de ser reutilizados y posteriormente adaptados a un entorno de aprendizaje (Kelty, Burrus, & Baraniuk, 2008). La reutilización, ha permitido tener mayor acogida dentro del ámbito académico, debido a la disminución de costes de producción para el desarrollo de materiales educativos. Esto, facilitaría las actividades de un profesor para desarrollar actividades con una participación activa, constructiva y cooperativa por parte de sus estudiantes (Howland, Jonassen, & Marra, 2012), y por otro, ser una respuesta para las instituciones académicas sobre la amplia demanda de recursos educativos abiertos en nuestra sociedad actual (Wiley & Hilton, 2009). Por lo tanto, la esencia del recurso digital, ha dado lugar al desarrollo de objetos de aprendizaje, sin embargo los objetos de aprendizaje van más allá del carácter abierto, presentando una serie de atributos de carácter tecnológico para propósitos

de creación y uso; algunos de ellos orientados hacia la modularidad y reutilización (N Friesen, 2009).

Al objeto de aprendizaje se le han asignado una serie de atributos asociados a sus características de uso. Una de las características que mejor lo describe, se encuentra relacionado al concepto de unidades independientes (Sánchez & Sicilia, 2005), donde encontramos propiedades asociadas a la reutilización sobre distintos entornos tecnológicos, y contextos de aprendizaje (Kelty, et al., 2008; Wiley, 2002a). Internacionalmente se han establecido normas para permitir su reutilización e interoperabilidad. Estándares como (Dublin Core^[1], ADL-SCORM^[2], IEEE-LOM^[3], Simple Knowledge Organization System-SKOS^[4], entre otros) y conjuntos de especificaciones (IMS del Global Learning Consortium^[5]), se han definido para que un objeto de aprendizaje disponga de una serie de descripciones, con el fin de ofrecer una mayor flexibilidad para ser implementado en distintas plataformas; evitando así, problemas de interoperabilidad (Aroyo et al., 2006; Baker, 2006; McGreal, 2008; Sanz, Dodero, & Sánchez, 2009; Sicilia & Garcia, 2003).

A partir de este panorama, podemos complementar que un recurso digital por sí solo carece de características que facilitan su acceso, uso y posterior reutilización. Por lo tanto para que un recurso digital pueda considerarse un objeto de aprendizaje, este debe contener, por lo menos, algunas de las siguientes características: i) diseño educativo e instruccional para su creación (Wiley, 2002b); ii) componentes pedagógicos para su uso (Polsani, 2006); iii) conjunto de descripciones denominados metadatos, los cuales permitirán conseguir cierto grado de interoperabilidad, accesibilidad y reutilización entre diferentes sistemas (N Friesen, 2009; Sanz, et al., 2009).

De acuerdo a los propósitos de este trabajo, se abordará la definición de objeto de aprendizaje desde la perspectiva ofrecida por Sánchez (2005):

“Unidad didáctica, independiente, autocontenida y perdurable, predispuesta para su reutilización en diversos contextos educativos mediante la inclusión de información autodescriptiva en forma de metadatos en formato digital estandarizados”.

A partir de esta definición, uno de los intereses de la presente investigación se orienta al uso de estos *metadatos*, con el propósito de definir criterios de búsqueda y acceso a material educativo a partir de la implementación de técnicas de visualización sobre repositorios digitales educativos.

1.1.2 Metadatos de objetos de aprendizaje

Los metadatos son unidades de datos cuya finalidad es ofrecer una descripción del objeto de aprendizaje. En el ámbito educativo, facilitan la construcción de un recurso digital y determinan los elementos apropiados para ciertas necesidades de aprendizaje. En el área de los objetos de aprendizaje, el uso de metadatos es ampliamente aceptado como medio para aumentar su

calidad y reusabilidad (Sicilia & Garcia, 2003). Por lo tanto, los metadatos se usan para describir los objetos de aprendizaje con el fin de facilitar procesos de búsqueda y recuperación de información (J. Fernández, 2001; Pastor, 2009). Son un requisito esencial para obtener resultados de búsqueda relevantes y satisfacer las necesidades de búsqueda de cualquier usuario (Cechinel, Sánchez, & Sicilia, 2009).

La búsqueda de un recurso digital, en particular, depende de la calidad de estos metadatos. La calidad de los metadatos es esencial en la búsqueda de recursos digitales; sin ellos, los resultados de búsqueda son pobres e ineficientes, dada la poca relevancia del gran número de recursos digitales obtenidos en una consulta (Cechinel, et al., 2009). El propósito del estudio de metadatos se orienta hacia la evaluación de la calidad de los recursos digitales con el fin de proporcionar niveles mínimos de aprendizaje (Hughes, 2005). Por lo tanto, la calidad de los metadatos facilita el desarrollo de repositorios para mejorar: i) actividades de gestión e indexación de materiales educativos (Ochoa, Cardinaels, Meire, & Duval, 2005; Stuckenschmidt, Vdovjak, Houben, & Broekstra, 2004); ii) la reusabilidad de materiales educativos en repositorios abiertos (Cervera, López, Fernández, & Sánchez, 2009; Sanz, et al., 2009); iii) la ubicación de material relevante mediante mecanismos de valoración de recursos digitales (Cechinel, Silva, Ochoa, Sánchez, & Sicilia, 2012; Ochoa & Duval, 2008).

La creación y gestión de metadatos es un proceso que puede realizarse a lo largo de todo el ciclo de vida de un recurso digital. Las descripciones de estos metadatos se pueden alojar de dos maneras: i) pueden residir dentro del mismo lugar de almacenamiento del recurso digital, mediante la vinculación a un conjunto de metadatos; ii) de manera externa mediante rigurosos sistemas de almacenamiento de metadatos, donde se incluye la referencia del objeto de aprendizaje. Para llevar a cabo cualquiera de las dos prácticas, es habitual contar con la participación de un repositorio digital.

1.1.3 Repositorios digitales educativos

Los repositorios son recursos informáticos que permiten el almacenamiento y gestión de objetos de aprendizaje. En el ámbito educativo han tenido gran acogida debido al gran número de recursos digitales que se comparten abiertamente sobre Internet. Los repositorios digitales educativos presentan dos modalidades de operación: i) los que permiten el almacenamiento de objetos de aprendizaje junto con sus metadatos, y ii) los que permiten el almacenamiento únicamente de los metadatos. Este último modelo es el que presenta un mayor crecimiento, dado que muy pocos repositorios permiten el almacenamiento de objetos de aprendizaje (Ochoa & Duval, 2009).

El diseño de repositorios digitales educativos no sólo ofrece la posibilidad de gestionar objetos de aprendizaje, sino también satisfacer necesidades de proveedores de contenidos que deseen exponer sus recursos educativos, (p.e. ARIADNE^[6], Organic.Lingua^[7], MERLOT^[8]) y para una comunidad de desarrolladores que deseen construir o adaptar de manera sencilla algunas de las herramientas que usan, a partir de servicios expuestos por el repositorio

(Santos, Ochoa, Parra, & Duval, 2011). Actualmente la mayoría de repositorios educativos cuentan con esquemas de representación formal de conocimiento, a partir del uso de tesauros y ontologías. Estos esquemas permiten incorporar semántica computacional para realizar actividades de clasificación, seguimiento, reutilización y búsqueda de objetos de aprendizaje.

Existen repositorios digitales orientados a exponer recursos de proveedores externos, con el objetivo de centralizar procesos de búsqueda a través de un único punto de acceso. Un claro ejemplo, es el proyecto Europeana^[9], una biblioteca digital encargada de recopilar y poner a disposición la mayor cantidad de recursos digitales de patrimonio cultural europeo. Estos recursos se encuentran físicamente almacenados en repositorios externos; por lo tanto, Europeana se encarga de recopilarlos, catalogarlos y organizarlos para desplegarlos a través de sus mecanismos de búsqueda. Otro ejemplo de este tipo de repositorios es MACE^[10], un proyecto orientado a ofrecer acceso a más de 150.000 recursos digitales asociados con el área del “diseño y arquitectura”. Define una serie de alternativas de búsquedas a partir de palabras claves, ubicación geográfica, competencias y área social (Y. Lin, Ahn, Brusilovsky, He, & Real, 2010; Stefaner et al., 2007). A pesar de ser repositorios digitales no educativos, presentan un gran número de recursos digitales que permiten ser candidatas para el reuso en entornos de aprendizaje.

El continuo desarrollo de propuestas e iniciativas orientadas al almacenamiento y gestión de recursos digitales, ha despertado el interés por parte de una comunidad de investigadores, orientados a estudiar la utilidad de este tipo de herramientas (Jeng, 2009; Kim & Kim, 2008; Qing & Ruhua, 2008; Tsakonas & Papatheodorou, 2006, 2008). Bajo esta perspectiva, a continuación se centra sobre una de las líneas encargadas de realizar este tipo de análisis, conocida con el nombre de usabilidad.

1.1.4 Usabilidad

La usabilidad es un concepto asociado a la facilidad de uso de un producto, aplicación o herramienta informática. A pesar de ser una palabra que no ha sido aceptado por la Real academia de la Lengua, es un término ampliamente utilizado en varios ámbitos. Uno de ellos, se encuentra orientado al análisis de factores asociados al funcionamiento de aplicaciones informáticas (Nielsen, 1994b, 2003). Varios autores coinciden en que la usabilidad es un factor determinante que permite medir la calidad de un producto informático y por tanto definir su éxito o fracaso (Perallos, 2007). Por su lado autores como Nielsen & Hackos (1994), definen que la usabilidad es un factor crítico para que el sistema alcance su objetivo. Existen varios enfoques y metodologías que permiten definir criterios de evaluación de la usabilidad, de acuerdo a un dominio de aplicación específico. Algunos de ellos orientados a la evaluación de sitios de comercio electrónico (Nielsen, Snyder, Molich, & Farrell, 2001), aplicaciones educativas (Alva et al., 2010), y otros, orientados sobre dominios de conocimiento general (Granollers, 2004; Hassan, Martín, & Iazza, 2004; Nielsen, 1994a, 2003; Vora, 1998).

Desde el punto de vista normativo, también existen varios estándares sobre usabilidad. Resaltamos la definición dada por la ISO/IEC 9126 (9126 ISO, 2001), donde se define la usabilidad como:

“la capacidad de un software de ser comprendido, aprendido, usado y atractivo para el usuario, en condiciones específicas de uso”.

Estos atributos, favorecen en cierta medida la calidad del software mediante la aplicación de características asociadas al diseño.

La usabilidad no se puede estudiar de forma aislada, por lo tanto es importante analizarla en relación con la utilidad de la aplicación, es decir, con el grado en que la aplicación pueda adaptarse a las exigencias de un usuario (Hassan, et al., 2004; Norman, 2005). Autores como Dillon & Morris (1999) definen la usabilidad como *“el grado en que el usuario puede explotar la utilidad”*. Es una característica de peso, dada la relevancia que se le da al usuario al experimentar y enriquecer el uso de una aplicación. En este sentido, tanto la usabilidad como la utilidad son factores que no se deben considerar de forma aislada para el análisis y evaluación de aplicaciones. Sin embargo, como lo menciona (Norman, 2005), en muchos casos será necesario que los usuarios logren adaptarse a la herramienta o producto, y conducirlos a un proceso que les permita identificar criterios para llevar a cabo una valoración sobre su usabilidad y utilidad.

1.2 Motivación

Uno de los mayores desafíos que enfrentan los repositorios de datos especializados, es la definición de métodos y estrategias que permitan al usuario facilitar el acceso a recursos digitales. Por lo tanto, existe la necesidad de definir estrategias de búsqueda adecuadas, donde se aborde la definición de esquemas de representación de conocimiento (ontologías y tesauros), el diseño y calidad de los metadatos y el desarrollo de interfaces flexibles. La identificación de las mejores estrategias permitirían, en cierta medida, mejorar las condiciones de acceso a los recursos digitales, los cuales dentro de un escenario educativo, son condiciones necesarias para crear un entorno de trabajo colaborativo con el fin de facilitar procesos para encontrar material educativo, y categóricamente cubrir áreas de conocimiento de interés.

Actualmente la mayoría de repositorios digitales incluyen esquemas de representación de conocimiento (ontologías, tesauros, etc.) con el propósito de ofrecer a creadores y expertos alternativas para gestionar, clasificar y organizar objetos de aprendizaje dentro de un repositorio; así como facilitar el uso del repositorio para ayudar a los usuarios a localizar objetos de aprendizaje, desplegar información complementaria, identificar sus relaciones, categorías, áreas de conocimiento, etc. Sin embargo, estos esquemas de clasificación a veces no son utilizados y aprovechados en la medida que sería deseable ya que: i) los usuarios no están familiarizados con los complejos sistemas de representación de conocimiento, que los

expertos y creadores de repositorios utilizan para clasificar los recursos, y ii) las interfaces de usuario actuales no siempre facilitan las tareas esenciales de los usuarios que visitan el repositorio: búsqueda y localización de recursos digitales. La motivación de esta investigación se centra en el segundo factor, presentando un análisis de las deficiencias de usabilidad y utilidad de las actividades de búsqueda y navegación mediante el diseño e implementación de interfaces gráficas a partir del uso de técnicas de visualización.

A veces las interfaces de búsqueda, pueden ser más un problema para los usuarios de un repositorio que buscan los recursos digitales adecuados. En consecuencia, el presente trabajo investiga si a través de técnicas de visualización y metodologías de Interacción Persona Ordenador (IPO), podemos ayudar a los creadores de repositorios digitales a ofrecer mejores servicios a sus usuarios con el fin de: i) localizar materiales de una manera más eficaz y precisa dentro de repositorios de recursos digitales, ii) ayudar a los usuarios a facilitar la ubicación de los materiales de acuerdo a una estructura temática o área de conocimiento, iii) identificar la cobertura temática de un área de conocimiento y un tema específico en función del número de recursos digitales; y, por último iv) identificar las interfaces más eficaces en función de los criterios de búsqueda, para llevar a cabo procesos de exploración y búsqueda de recursos digitales.

Uno de los propósitos que persigue esta investigación es explorar posibles formas de desplegar información dentro de una colección de recursos digitales alojados en un repositorio, y analizar sus diferencias con el fin de llegar a conclusiones que permitan identificar ventajas y desventajas de cada una de ellas. Para identificar principios del diseño de interfaces, utilizaremos como estrategias diversas técnicas de visualización de información (Herman, Melancon, & Marshall, 2000), que de acuerdo con propósitos de este estudio, hemos seleccionado a partir de características asociadas con: i) estructura de representación gráfica (Herman, et al., 2000), ii) estrategias de agrupamiento temático (Merčun, Žumer, & Aalberg, 2012), y finalmente iii) representaciones jerárquicas para despliegue de categorías (Schulz, Hadlak, & Schumann, 2011; Shneiderman, Feldman, Rose, & Grau, 2000). En base a estos criterios de selección, se han desarrollado ocho interfaces de visualización, visualmente clasificadas en cuatro estrategias: *radiales*, *hiperbólicas*, *árboles* y por *categorías*. En el Capítulo 5 se llevará a cabo un estudio de usabilidad para evaluar las capacidades de percepción visual de cada una de las ocho interfaces. Estas capacidades se encuentran asociadas con los siguientes atributos: *atención*, *retención* y *comprensión* de temas relacionados con un área de conocimiento. En la misma dirección, la capacidad de interacción es otra de las características dentro del área de Interacción Persona Ordenador que permite, en cierta medida, obtener la impresión subjetiva y objetiva de un usuario. Este estudio se ampliará con mayor detalle en el Capítulo 4 de esta memoria, a partir de la ejecución de un caso de estudio, donde se parte de la base de un esquema de representación de conocimiento mediante el uso de un tesoro de Arte y Arquitectura (AAT)^[11], y un conjunto de recursos digitales de la biblioteca digital Europea^[9].

Por otro lado, y de manera complementaria, el éxito de las estrategias de búsqueda, depende de lo bien definido que se encuentren los metadatos para describir un objeto de aprendizaje.

Por tal motivo, es pertinente estudiar la calidad de los mismos. Esta actividad se ampliará en mayor detalle en la Sección 4.2.4, asociado con el análisis de calidad de metadatos a partir de una colección de recursos digitales extraídos de la biblioteca digital Europea como caso de estudio. Los metadatos explorados a partir de estos recursos digitales fueron extraídos de acuerdo a temáticas relacionadas de un vocabulario controlado de términos definidos por el área de conocimiento de “*Styles and Periods*” del tesauro AAT, con el fin de lograr asociarlos a un esquema de representación de conocimiento. Por lo tanto, este estudio complementario, pretende analizar la calidad de los metadatos de los recursos digitales extraídos de Europea, con el propósito de identificar criterios de búsquedas alternativos para llevar a cabo la implementación de las interfaces visuales, tales como: búsqueda del término a partir del formato del recurso digital (audio, video, imagen), a partir de su idioma, proveedor de contenidos, etc.

Finalmente, uno de los propósitos que se persigue con este trabajo de investigación es ofrecer una alternativa diferente a las búsquedas tradicionales que ofrecen los repositorios para localizar recursos digitales, mediante la implementación de un escenario real de trabajo. Esto es, el planteamiento de estrategias de búsquedas visuales mediante técnicas de visualización a partir del uso de esquemas de representación de conocimiento (tesauro AAT), para facilitar el acceso a objetos de aprendizaje (recursos digitales de Europea) a través de búsquedas navegacionales de acuerdo a un área de conocimiento (temática). Estas categorías permitirían definir una serie de criterios de búsqueda visuales (p.e. idioma, proveedor, derechos de autor, etc.), los cuales se encuentran descritos a partir de sus metadatos.

1.3 Objetivos y contribuciones

El **objetivo principal** de esta investigación es estudiar alternativas de búsqueda mediante el uso de técnicas de visualización, con el objeto de mejorar el acceso a objetos de aprendizaje sobre repositorios digitales que utilizan esquemas de representación de conocimiento.

Para alcanzar este objetivo general, se plantean los siguientes **objetivos específicos**:

(O1) Estudiar el uso de esquemas de representación de conocimiento en repositorios digitales y las estrategias de búsquedas asociadas.

(O2) Investigar e identificar las diferentes técnicas de visualización de información que permitan desplegar esquemas de representación de conocimiento en repositorios digitales.

(O3) Desarrollar un prototipo que permita integrar técnicas de visualización y búsqueda de objetos de aprendizaje en repositorios digitales a partir de criterios definidos por sus metadatos.

(O3.1) Diseñar y/o adaptar interfaces mediante técnicas de visualización a partir de esquemas de representación de conocimiento para desplegar información asociada a un área de conocimiento.

(O3.2) Desarrollar una herramienta que permita realizar extracción de recursos sobre un repositorio digital a partir de un área de conocimiento definido por un esquema de representación de conocimiento.

(O3.3) Analizar la calidad de los metadatos de los recursos digitales extraídos en (O3.2) e identificar criterios de búsquedas complementarios sobre cada una de las interfaces gráficas bajo estudio.

(O.4) Realizar un estudio de usabilidad de las interfaces visuales diseñadas para identificar las ventajas y deficiencias de cada una de las técnicas de visualización bajo estudio.

1.3.1 Contribuciones

Una de las aportaciones centrales de esta investigación es ofrecer una alternativa de búsqueda visual a partir de técnicas de visualización para el acceso a una colección de recursos digitales sobre repositorios académicos. El propósito es plantear una serie de recomendaciones para los creadores de repositorios digitales, antes de llevar a cabo un desarrollo formal de interfaces gráficas mediante el uso de técnicas de visualización. Esto disminuirá, previsiblemente el tiempo, esfuerzo y factores económicos en la implementación de este tipo de alternativas dentro de repositorios digitales. Entre las aportaciones complementarias que han surgido a partir de esta contribución principal, podemos destacar las siguientes:

1. Análisis de calidad de metadatos de recursos digitales de biblioteca digital Europea, donde se han identificado algunas de las limitaciones y se plantearon recomendaciones sobre el modelo de representación de datos para la búsqueda de recursos digitales a partir de la cobertura temática (Gaona, Feroso, & Sánchez, n.p.).
2. Aplicación de principios de Interacción Persona Ordenador (IPO) para plantear análisis de usabilidad de las interfaces diseñadas a partir de las técnicas de visualización (Gaona, et al., 2014).
3. La definición de una serie de recomendaciones para el desarrollo de interfaces visuales a partir de técnicas de visualización sobre repositorios digitales para el acceso a grandes volúmenes de recursos digitales.
4. Desarrollo de un framework e integración de métodos de búsquedas tradicionales, con métodos de búsqueda visuales a partir de técnicas de visualización. (Gaona, Sánchez, & Montenegro, 2014).
5. Integración de métodos de valoración de recursos digitales dentro de interfaces de navegación. Método complementario para desplegar el listado de los recursos digitales relevantes a partir de la colección de recursos digitales identificados inicialmente por área de conocimiento.

1.4 Método de trabajo

A continuación se presenta la estrategia de trabajo usada para llevar a cabo la planificación de cada una de las actividades.

1.4.1 Estudio del estado de la cuestión

Se llevó a cabo un estudio preliminar de exploración de recursos digitales tomando como caso de estudio la biblioteca digital Europea. Usando esta biblioteca, se llevó a cabo un proceso de análisis de cobertura temática para identificar el número de recursos digitales de Europea, asociados a un área de conocimiento específico (Gaona, et al., n.p.). El concepto de cobertura temática, es una adaptación que se tomó a partir de definiciones realizadas por (Codina, 2000) y Whitehall (1992), con el propósito de analizar en qué medida Europea cubre ciertas áreas de conocimiento a partir de una rama de conocimiento definidos por un tesauro de Arte y Arquitectura AAT (Soergel, 1995; Trust, 1988). Este estudio permitió identificar dos líneas de análisis y exploración. Por un lado el análisis de calidad de metadatos de los recursos digitales extraídos, y por el otro, un análisis y exploración bibliográfica para llevar a cabo el desarrollo e implementación de interfaces visuales a partir de técnicas de visualización de información para facilitar el acceso a recursos digitales. Finalmente esta última línea de análisis, sumado a problemas identificados en estudios realizados sobre los repositorios académicos Organic.Edunet y VOA3R (Martín, et. al. 2013), permitieron vislumbrar el tema central de investigación, al identificar deficiencias en el uso de interfaces de búsquedas sobre repositorios digitales a partir de sus esquemas de representación de conocimiento. Adicional a ello, se parte del hecho de la escasa literatura relacionada con las búsquedas visuales mediante estrategias de visualización taxonómica en repositorios digitales y estudios de usabilidad relacionados.

1.4.2 Identificación y descripción del problema

El problema se plantea, a partir de la fase preliminar de exploración de recursos digitales de la biblioteca digital Europea (Gaona, et al., n.p.). Como resultado de este proceso, se identificó que una gran mayoría de recursos digitales no se encontraban relacionados con el área de conocimiento definido por el tesauro AAT. Esto condujo a realizar un análisis de calidad de sus metadatos, encontrando problemas asociados a nivel de completitud, exactitud, redundancia e inconsistencias.

Por otro lado, mediante revisión bibliográfica se encontró que parte de las carencias significativas del uso repositorios digitales, se encontraban asociados al grado de usabilidad de las interfaces. Dentro de la revisión bibliográfica, muy pocos estudios tratan el grado de usabilidad de interfaces en repositorios digitales, además de la escasa bibliografía asociada con el diseño de interfaces mediante estrategias de visualización taxonómica. Estos elementos, junto con los resultados de búsqueda no significativos identificados en la fase preliminar de este estudio, fueron algunos de los problemas que dieron lugar a la actual investigación.

1.4.3 Definición del modelo

A partir de la clasificación definida por el esquema de representación de conocimiento del tesoro AAT, se definió un conjunto de términos mediante la aplicación de una serie de criterios taxonómicos como el nivel de profundidad, el número de términos asociados por categoría, la relación de términos padre por categoría, y el balanceo de carga de cada una de las ramas o categorías seleccionadas. Este modelo, fue la base para llevar a cabo un proceso de transformación de datos y mapeo de los recursos digitales extraídos a partir de la biblioteca digital Europea. Una vez con este esquema de representación de conocimiento, se llevó a cabo su implementación a partir de las ocho técnicas de visualización, previamente identificadas.

1.4.4 Análisis de calidad de metadatos

De manera simultánea a la definición del modelo de representación taxonómica, se llevó a cabo un análisis de la calidad de metadatos extraídos a partir del proceso de exploración de recursos digitales de Europea (Gaona, et al., n.p.). El propósito central de este estudio, era evaluar en qué medida Europea cubría ciertas áreas de conocimiento a partir de un conjunto de términos definidos previamente a partir del tesoro de Arte y Arquitectura AAT. Esta cobertura, permitió identificar, entre otras cosas, cuáles eran los metadatos que ofrecían un mayor grado de completitud con el propósito de tomarlos como criterios de búsqueda de recursos digitales a partir de técnicas de visualización. Sin embargo, a lo largo del estudio, se identificaron una serie de deficiencias en la definición de los metadatos asociados con criterios de calidad de metadatos tales como: *completitud*, *exactitud*, *redundancia* y *ambigüedad* en la descripción sobre los recursos digitales explorados. Este tipo de resultados, conducen finalmente a resultados de búsquedas ambiguos, que no se encuentran asociados a un área de conocimiento seleccionado.

1.4.5 Definición de un framework de visualización

Se llevó a cabo el desarrollo de un framework mediante el cual se definió la integración de las interfaces visuales implementadas para llevar a cabo tres estrategias de búsqueda, a saber: i) búsqueda navegacional, mediante la exploración por categorías a partir de cada una de las técnicas de visualización implementadas en las interfaces; ii) búsqueda textual, mediante la integración de un componente de búsqueda que permite autocompletar el término seleccionado, y ubicar el término dentro de la estructura de navegación gráfica; y finalmente iii) búsqueda visual, mediante la distribución del número de recursos digitales asociados al término, a partir de criterios de interés definidos por el idioma, tipo de recurso digital, derechos de autor, etc. Este framework permitió la navegación de los recursos digitales asociados a un área de conocimiento del término seleccionado y la integración de un esquema de representación de conocimiento definido por el tesoro AAT.

1.4.6 Estudio de usabilidad

A partir del uso de las interfaces visuales implementadas en el framework, se lleva a cabo un estudio de usabilidad mediante principios básicos asociados a Interacción Persona Ordenador. Por un lado se llevó a cabo la evaluación de factores de percepción visual, impresión subjetiva y otros criterios asociadas con la experiencia de las interfaces por parte de los usuarios (evaluación subjetiva). Por otro lado se presentó un estudio de la interacción de los usuarios mediante la aplicación de un caso de estudio, para realizar actividades de búsqueda de términos mediante procesos de navegación y exploración (evaluación objetiva).

1.4.7 Análisis de resultados

Se llevó a cabo la interpretación de los resultados a partir de criterios de evaluación definidos previamente para medir el grado de usabilidad. Para llevar a cabo la evaluación se tuvieron en cuenta dos resultados: i) el nivel de percepción visual del usuario (evaluación subjetiva), y ii) la interacción y resultados de búsqueda obtenidos a partir de un caso de estudio (evaluación objetiva).

1.4.8 Formulación de conclusiones

Finalmente se exponen los resultados de los estudios. Por un lado el análisis de metadatos de recursos digitales de Europeana, los cuáles permitieron identificar algunos criterios de búsqueda para la implementación sobre interfaces gráficas. Por otro lado, los concernientes a nuestro objeto principal de estudio de interfaces basadas en técnicas de visualización, donde se plantean algunas directrices, recomendaciones y sugerencias para el desarrollo de interfaces visuales mediante técnicas de visualización para el acceso a una colección de recursos digitales; y posteriormente se definen algunas líneas de trabajo futuras.

1.5 Estructura del documento

Esta memoria de tesis doctoral se encuentra dividida en los siguientes capítulos:

1. El primer capítulo presenta la introducción del tema de interés, desglosando de manera general el contexto de los temas que se abordaran a lo largo de esta investigación. A continuación se expone la motivación que ha dado lugar al desarrollo del presente tema de investigación, así como sus objetivos. Finalmente se presenta el método de trabajo y la planificación que se llevó a cabo para abordar el tema de trabajo.
2. En el segundo capítulo se presenta el Estado de la cuestión, donde se exponen de manera detallada cada uno de los temas del estudio asociados con esquemas de representación de conocimiento, la importancia de las taxonomías y las jerarquías en la

clasificación de los objetos de aprendizaje, y las estrategias de acceso implementadas por repositorios. Seguidamente se aborda la importancia de las estrategias de visualización de información sobre las bibliotecas y repositorios digitales. A continuación se presentan las técnicas de visualización estudiadas. Finalmente se analiza la importancia de la usabilidad aplicada sobre la navegación de una colección de recursos digitales, sus implicaciones y los trabajos relacionados en el área.

3. El capítulo tercero, se realiza el planteamiento del problema, partiendo de las deficiencias de los métodos de búsqueda y acceso sobre objetos de aprendizaje, seguido de las limitaciones de los esquemas de representación de conocimiento usados en repositorios, y finalmente exponiendo las limitaciones y carencias presentadas en la definición de metadatos, tomando para ello como caso de estudio, a recursos digitales de la biblioteca digital Europea.
4. En el cuarto capítulo se presenta el diseño de una solución a partir de la definición de un caso de estudio. En un primer apartado se define la metodología y las estrategias que se llevaron a cabo para plantear el escenario de trabajo. A continuación la fase de exploración y extracción de recursos digitales de la biblioteca digital Europea. Finalmente se presenta la manera en que se llevó a cabo la definición del esquema de representación de conocimiento, seguido del diseño e implementación de las interfaces de visualización.
5. En el capítulo quinto se expone la metodología que se llevó a cabo para realizar el estudio de usabilidad, presentando en primera instancia la estrategia de selección de participantes, y los criterios de selección de las pruebas de usabilidad definidas en dos fases, a saber: percepción e interacción.
6. El capítulo sexto se presentan los análisis de resultados obtenidos en las pruebas de usabilidad de las interfaces implementadas. De la misma forma se presenta el análisis de calidad de metadatos obtenidos en el proceso de extracción de recursos digitales realizado en el capítulo cuarto. Finalmente se presentan las discusiones de los resultados obtenidos y se exponen una serie de propuestas, producto de los resultados y experiencias obtenidas a lo largo del presente estudio, con el propósito de ser útiles a creadores de repositorios a la hora de desarrollar e implementar interfaces a partir de técnicas de visualización.
7. Finalmente, en el capítulo séptimo se exponen las conclusiones, algunas recomendaciones sobre el desarrollo de interfaces a partir de técnicas de visualización, se apuntan trabajos futuros de investigación, y se mencionan las publicaciones derivadas a lo largo del presente estudio.

Pie de páginas:

- [1] Disponible online en: <http://dublincore.org/>, (último acceso 5 Mayo 2014)
- [2] Disponible online en: <http://www.adlnet.gov/scorm>, (último acceso 5 Mayo 2014)
- [3] Disponible online en: <http://ltsc.ieee.org/wg12/>, (último acceso 5 Mayo 2014)
- [4] Disponible online en: <http://www.w3.org/2004/02/skos/>, (último acceso 5 Mayo 2014)
- [5] Disponible online en: <http://www.imsglobal.org/>, (último acceso 5 Mayo 2014)
- [6] Disponible online en: <http://www.ariadne-eu.org/>, (último acceso 5 Mayo 2014)
- [7] Disponible online en: <http://www.organic-lingua.eu/es>, (último acceso 5 Mayo 2014)
- [8] Disponible online en: <http://www.merlot.org/>, (último acceso 5 Mayo 2014)
- [9] Disponible online en: <http://europeana.eu/>, (último acceso 5 Mayo 2014)
- [10] Disponible online en: <http://portal.mace-project.eu/>, (último acceso 5 Mayo 2014)
- [11] Disponible online en: <http://www.getty.edu/research/tools/vocabularies/aat/>, (último acceso 5 Mayo 2014)

Capítulo 2

Estado de la Cuestión

El propósito de este capítulo es presentar un panorama general sobre los temas de interés asociados a nuestro proyecto. Por un lado presentamos la importancia de los esquemas de representación de conocimiento, como estructuras que facilitan la clasificación y organización de contenidos en repositorios digitales. A continuación se resaltan los instrumentos más comunes para acceder al conocimiento dentro del ámbito de los repositorios digitales a través del uso de ontologías y tesauros, y las características de las taxonomías como instrumentos esenciales para llevar a cabo este tipo de clasificaciones. Posteriormente, se aborda el tema de la visualización y las estrategias de acceso a recursos digitales en repositorios. Finalmente se expone la importancia de la usabilidad, el uso de taxonomías y esquemas de representación de conocimiento para el acceso a recursos digitales, seguido de una revisión bibliográfica de los trabajos más representativos asociados al tema de estudio.

Capítulo 2: Estado de la Cuestión

*“Cada día sabemos más y entendemos menos”
(Albert Einstein)*

2.1. Esquemas de representación de conocimiento

En la literatura se encuentra una variada lista de definiciones asociados a *un esquema de representación de conocimiento*, partiendo de la filosofía, el razonamiento humano, las ciencias cognitivas y la inteligencia artificial. Sin embargo, para los propósitos de esta investigación, se desglosarán cada uno de los conceptos que permitan abordar su significado con el propósito de comprender su valor e importancia en el área de los *repositorios digitales*. Para ello, en primera instancia se aborda el concepto de *conocimiento*; término muy usado a lo largo del desarrollo tecnológico, para describir la importancia de la gestión de información como unidades de valor agregado para llevar a cabo actividades que impactan sobre nuestra sociedad y economía. No obstante, de acuerdo a los propósitos de esta investigación, se orientará sobre el área de las ciencias de la computación, específicamente en la línea de la inteligencia artificial, área que ofrece un acercamiento más claro de este concepto.

A la fecha, el estudio del conocimiento humano y su representación en el área de la computación, presenta todavía un concepto muy centralizado y ampliamente difundido en la rama de la inteligencia artificial (Gašević, Djurić, & Devedzic, 2009). En este sentido, el conocimiento desde la perspectiva de la inteligencia artificial se define como la información de un dominio que puede ser utilizado para resolver problemas sobre el mismo dominio (Brachman & Schmolze, 1985). Por lo tanto, para resolver grandes problemas, se requiere de un mayor conocimiento, y este conocimiento debe estar representado de alguna manera sobre un computador. Para resolver estos problemas, se diseñan aplicaciones que permiten definir la manera en que es representado el conocimiento (Gero, 1990). Un ejemplo de estas aplicaciones, son los agentes de software; programas diseñados para llevar a cabo una actividad en particular, y actuar bajo ciertas reglas o patrones de diseño de acuerdo a un conjunto de acciones de aprendizaje o razonamiento (Genesereth & Ketchpel, 1994), a partir de un *esquema de representación de conocimiento*. En este sentido, un *esquema de representación* especifica la forma en que estos conocimientos pueden ser usados por un agente de software.

De acuerdo a la perspectiva ofrecida por Durkin (1994); el *conocimiento* lo podemos asociar a partir de un área temática donde se incluyen conceptos y relaciones asociados a la misma. Esta aproximación nos permite identificar dos elementos, a saber: i) *área temática*, el cual entendemos como un dominio o conjunto de temas y conceptos de interés; y por otro lado ii) las *relaciones*, quienes son las encargadas de mantener una asociación con aquellas áreas de interés entre sí, a partir de una afinidad temática, jerárquica, histórica, etc., según sea el caso. A partir de esta apreciación, Durkin (1994) plantea una serie de clasificaciones que describen la

manera en que podemos representar el conocimiento humano. En la Figura 1, se representa una adaptación de su propuesta:

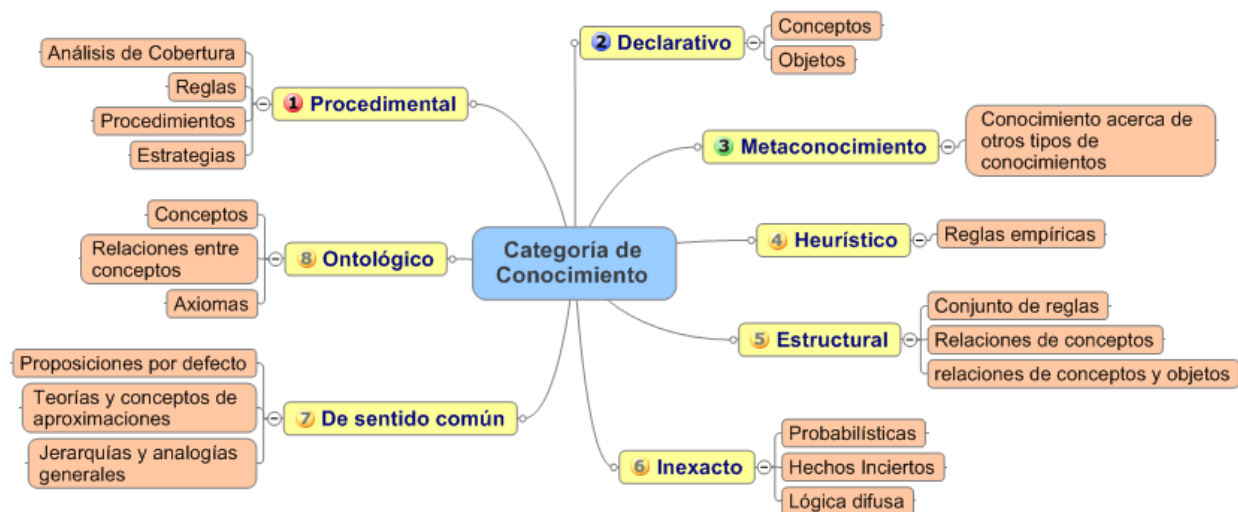


Figura 1. Tipos de conocimiento humano y su representación (Durkin, 1994)

De manera general el *conocimiento procedimental* se relaciona con las maneras en que una persona debe actuar para resolver un problema en particular. El *conocimiento declarativo*, como su nombre lo indica, es un conocimiento que trata de describir lo que se conoce sobre un tema o problema. Por su parte el *metaconocimiento*, es usado para decidir qué otro conocimiento sería ideal para resolver un problema, y qué otros conocimientos serían irrelevantes de cara a resolver el problema. El *heurístico*, es un conocimiento que trata de incluir todas las reglas empíricas que guían a la resolución de un problema basado en la experiencia. El *conocimiento estructural*, es un conocimiento que se orienta a describir modelos mentales para llevar a cabo la organización de problemas y sus soluciones. El *conocimiento inexacto*, es caracterizado por la resolución de problemas, temas y situaciones que describe información imprecisa, incompleta, aleatoria o ambigua. El *conocimiento de sentido común*, es un término usado para denotar una amplia gama de conocimiento humano acerca del mundo que no se puede describir de forma precisa a través de teorías. Finalmente el *conocimiento ontológico*, es un complemento esencial del conocimiento acerca de un dominio específico, que describe categorías y cosas dentro del dominio, así como los términos que las personas usan para consultarlo.

De manera diferente al razonamiento humano, las computadoras por sí solas no poseen un mecanismo preciso que permita adquirir y representar conocimiento; para ello, necesitan de la supervisión de personas encargadas de almacenar conocimiento en la memoria de cada computadora. Por lo tanto, este tipo de actividades requiere de la supervisión continua de personas para decidir cómo representar el conocimiento, y de esta forma complementar el proceso para su posterior análisis y uso específico.

De lejos, la inteligencia artificial ha proporcionado una serie de estrategias y tipos de *representación de conocimiento* (Russell, Norvig, & Davis, 2010). Estas se encuentran

asociadas con métodos y estructuras para codificar un conocimiento en sistemas inteligentes, y determinar las relaciones existentes en su estructura de datos. Así, en un *esquema de representación* se pueden identificar estructuras de datos definidas a partir del uso de tablas, árboles, enlaces, grafos, etc. Cada estructura de datos poseen ventajas y deficiencias para representar diferentes tipos de conocimiento; por lo tanto, como lo indica Gasevic (2009), no existe un método de representación de conocimiento genérico que permita de manera estándar estructurar datos y trabajar sobre cualquier tipo de aplicación. Para realizar un recorrido por estas estructuras de representación de conocimiento, cada usuario lleva a cabo un proceso de navegación. Este proceso es esencial dentro del área de visualización de información (Martin Graham, Jessie Kennedy, & David Benyon, 2000). Por lo tanto es un concepto que se encuentra asociado a la manera en que un usuario procede a explorar y recorrer una estructura jerárquica definida como instrumento de representación taxonómica. Existen diferentes formas de llevar a cabo un proceso de navegación, las cuales se ampliarán en mayor detalle en la Sección 2.5.

Tomando como base las anteriores aproximaciones, se puede indicar que un *esquema de representación de conocimiento* dentro del ámbito de los repositorios digitales son:

“las diferentes formas o maneras que podemos usar para estructurar y representar datos, con el propósito de facilitar procesos de clasificación, organización y asociación de conceptos, a partir de un dominio de conocimiento o área temática previamente definida”.

Según esta perspectiva, podemos complementar que esta representación dentro de un entorno de aprendizaje facilitaría:

- La comprensión de una serie de temáticas asociadas con un área de conocimiento, a partir de un subconjunto de categorías relacionadas entre sí (Chrysafiadi & Virvou, 2013).
- La navegación de un dominio de conocimiento, a partir de una estructura de datos previamente definida (Shneiderman, et al., 2000; Z. Wang, Chaudhry, & Khoo, 2008).
- La búsqueda de recursos digitales, a partir de la descripción de sus metadatos mediante criterios de selección por áreas de conocimiento (Marchionini & White, 2010).
- La asociación de áreas temáticas afines, de acuerdo al nivel de relación que presente el esquema de representación de conocimiento de manera jerárquica (Uddin & Janecek, 2007).

Tomando como base este enfoque, en el siguiente apartado se presentan las formas más comunes de representación de conocimiento definidas para el diseño de repositorios digitales.

2.2. Formas de representar el conocimiento en repositorios digitales

La mayoría de aplicaciones informáticas emplean algún tipo de conocimiento, normalmente asociado como datos para su posterior aplicación. Este hecho, hace que la manipulación, gestión y administración de contenidos tenga una notable carga de trabajo, sea operativa, o actividades asociadas a la toma de decisiones. En algunos casos, el conocimiento se representa como datos en bruto que con frecuencia son almacenados en complejas estructuras de acuerdo a: secuencias lógicas, definición de reglas, árboles, grafos semánticos, y otras formas de representación (Flouris, Plexousakis, & Antoniou, 2003). A partir de estas estructuras se refleja formalmente una clasificación, y en el mejor de los casos, una relación semántica entre los datos, a nivel de dependencia, asociación, afinidad, etc.

La evolución de los repositorios digitales ha permitido centralizar procesos de gestión, administración y acceso a miles de recursos digitales. Sin embargo, el rápido crecimiento que ha tenido el desarrollo de materiales educativos ha generado un desafío sobre los creadores de repositorios digitales orientados a organizar, clasificar y gestionar los contenidos. Por lo tanto, una de las estrategias de mayor acogida para este tipo de necesidades a lo largo de los últimos años es el uso de instrumentos de clasificación a partir de *esquemas de representación de conocimiento*.

Cada técnica de representación de conocimiento requiere de una notación, la cual determina aspectos relacionados con el área temática, niveles de relación, vínculos y formas de asociación, entre otras. Para ser efectivo dentro de un ambiente de aprendizaje, un esquema de representación de conocimiento debe ser consistente y lo más realista para representar un área temática, y sus relaciones (Chrysafiadi & Virvou, 2013). Esto permitiría, entre otras cosas, aumentar o disminuir los conocimientos de un usuario que esté haciendo uso del mismo.

Para el diseño de repositorios digitales, los instrumentos usados para representar el conocimiento son: las taxonomías, las ontologías, los tesauros, además de los grafos, y los mapas mentales, entre otros. A continuación, se lleva a cabo un acercamiento a los tres primeros instrumentos, al ser *esquemas de representación de conocimiento* los más comunes para el diseño de repositorios digitales.

2.2.1 Taxonomías

Esta palabra viene del griego (τάξις) *taxis* = ordenación y (νομος) *nomia* = reglas, reglas de ordenación o clasificación. La definición de taxonomía es un concepto asociado a la rama de la biología, usado para realizar procesos sencillos de clasificación de organismos, plantas, géneros, reinos, formas, etc., donde varias especies se encuentran agrupadas de acuerdo a un conjunto de similitudes percibidas a partir de su forma, orden o clase (Stace, 1991). La taxonomía es también un concepto asociado a entornos digitales y al desarrollo de sitios Web

(Morante, 2003), donde proporciona una ruta de navegación estructurada sobre una colección de contenidos. Por lo general, el proceso de creación de una taxonomía se define a partir de un conjunto de datos relacionados entre sí. En el área de la visualización de información, esta definición da lugar a muchas alternativas taxonómicas para representar de manera jerárquica relaciones y niveles de profundidad (Graham & Kennedy, 2005). Para representar con exactitud esta multiplicidad de jerarquías taxonómicas, estudios previos han resaltado la importancia del desarrollo de modelos de datos, mediante el uso de sistemas de almacenamiento definidos a partir de estrategias como las bases de datos (Pullan, Watson, Kennedy, Raguenaud, & Hyam, 2000; Ytow, Morse, & Roberts, 2001; Zhong, Luo, Pramanik, & Beaman, 1999).

Asociadas al entorno digital, las taxonomías presentan un conjunto de finalidades orientadas a la mejora de procesos de navegación, desarrollo de sistemas de búsqueda basados en la exploración y la recuperación de información (Centelles, 2005). Estas taxonomías, han facilitado el proceso para gestionar una amplia base de conocimiento a través de nomenclaturas y lenguajes de representación; por lo cual, se presentan como instrumentos válidos para desplegar un amplio dominio de términos y sus relaciones.

Dentro del ámbito normativo, existen estándares que permiten definir y estructurar elementos que facilitan la creación de una taxonomía. Bajo esta perspectiva, la norma ANSI/NISO Z39.19 (ANSI/NISO, 2005), presenta cuatro tipos de clasificación sobre vocabularios controlados, a saber:

- *Lista*: Conjunto de palabras o frases desplegadas en series organizadas.
- *Anillo de sinónimos*: Conjunto de palabras o frases que se consideran equivalentes con el propósito de ser usadas para recuperar información.
- *Taxonomía*: Conjunto de palabras o frases organizadas, usadas para organizar información, y destinados principalmente para procesos de navegación.
- *Tesaurus*: Conjunto de palabras o frases con términos equivalentes y con palabras o frases ambiguas. Dentro de este conjunto de términos, se pueden incluir asociaciones de tipo padre/hijo y otro tipo de relaciones.

A partir de esta clasificación, podemos identificar que uno de los propósitos que persigue la definición de una taxonomía es facilitar tareas de clasificación y organización para llevar a cabo actividades de exploración y navegación dentro de entornos digitales. Desde este enfoque, a continuación abordamos los aspectos más relevantes del uso de taxonomías, a partir de la perspectiva ofrecida por (Gilchrist, Kibby, Mahon, & Libra, 2000):

- Ofrece una correlación de los diferentes lenguajes funcionales utilizados por una institución.
- Apoya mecanismos para navegar y acceder al capital intelectual de una institución.

- Proporciona herramientas de ayuda para navegar en un portal Web, etiquetar documentos y otros objetos de información, con el fin de apoyar la tarea de los motores de búsqueda y mapas de conocimiento.
- La taxonomía es una base de conocimiento en sí mismo.

En el área de los repositorios digitales, las taxonomías se relacionan con el uso de estructuras de representación jerárquicas para organizar y clasificar información compleja. Más que un instrumento, la taxonomía se considera como la base para la definición de un esquema de representación de conocimiento. En este sentido, los recursos digitales presentan un conjunto de relaciones que permiten clasificarlos mediante la definición de una serie de etiquetas a partir de términos asociados a su taxonomía. A estas etiquetas se les conoce con el nombre de metadatos, los cuales son encargados de describir de manera enriquecida un recurso digital. La Figura 2, presenta un ejemplo de la descripción de un recurso digital del repositorio digital MACE, donde se puede ver esta representación a partir de la definición de sus metadatos.

The screenshot displays the 'Content metadata' section of the MACE repository. It is divided into two main panels. The left panel shows a 3D architectural model of a building with a dome and a tower, labeled 'Model'. The right panel contains five numbered sections: (1) LANGUAGE: English; (2) RESOURCE TYPE: (2); (3) DESCRIPTION: Dal Co "F. Gehry opere e progetti" Electa p. 292; (4) Classification: Theoretical Concepts (model, building shape, estrangement), Relation with Context (estrangement), Styles, Periods and Trends (Pop); (5) Competencies: Architecture Competency Classification (selected), Engineering Competency Classification, Landscape Architecture, Urban Design, Bio-Climatic Design.

Figura 2. Descripción de un recurso digital a partir de sus metadatos - Ejemplo repositorio MACE

En el ejemplo de la figura 2, se puede ver un recurso digital y cinco descripciones de sus metadatos: 1) "language", 2) "resource type", 3) "Description", 4) "classification" y finalmente 5) "competencies". Estas descripciones, facilitan la clasificación y relaciones de un recurso digital asociado a un área de conocimiento. Por lo tanto, a partir de una taxonomía se puede definir la

estructura, o forma en que se representa un conjunto de conceptos. Estos conceptos se encuentran clasificados a partir de niveles de jerarquía, los cuales permitirán organizar el dominio de conocimiento que se pretende modelar.

2.2.2 Tesauros

La norma ISO 2788-1986 (ISO, 1985) define un tesauro como:

"un vocabulario controlado y dinámico, compuesto por términos que tienen entre ellos relaciones semánticas y genéricas y que se aplica a un dominio particular del conocimiento".

Normalmente los tesauros se encuentran asociados a un dominio de conocimiento. El uso de tesauros facilita procesos de indización y diseño de estructuras conceptuales (Arano & Codina, 2004), favorece el diseño de estructuras taxonómicas (Gilchrist, 2003), y simplifica el manejo y control de vocabulario (Nakayama, Hara, & Nishio, 2007). Presenta un amplio margen de uso en áreas de educación para la gestión de recursos digitales (A. Fernández, 2011) y documentación, para la clasificación y organización de materiales (L. Castillo, 2006), y se consideran en general instrumentos válidos para llevar a cabo actividades de recuperación de información (J. Fernández, 2001; Pastor, 2009).

Los tesauros se componen de unidades léxicas llamados términos preferidos (D) y términos no preferidos (ND). Los *términos preferidos (Descriptores)*, son unidades lingüísticas que expresan conceptos dentro de un área de conocimiento específico. Un concepto se expresa a partir de un único término, y ese término responde a un único concepto. Los *términos no preferidos (No Descriptores)*, son sinónimos de los términos preferidos (descriptores).

Dentro de la construcción de un tesauro, se definen una serie de relaciones que permiten manejar jerarquías, asociaciones o equivalencias:

- *Relaciones de jerarquía:* Expresa la subordinación entre dos términos en función de su significado. Por tanto una relación es jerárquica, cuando un término superior (genérico) engloba conceptualmente a otro (específico), generalmente conocido como relaciones de tipo Padre/Hijo o Género/Especie. De igual forma se pueden encontrar relaciones *poli jerárquicas*, donde cada hijo puede estar vinculado o asociado a varios padres.
- *Relaciones de asociación:* Es la que se establece entre términos próximos entre sí, que representan conceptos asociados por ideas, pero que no son equivalentes, ni existen entre ellos alguna relación de jerarquía.
- *Relaciones de equivalencia:* Son relaciones entre sinónimos o nombres que se refieren al mismo concepto.

El uso de tesauros en repositorios digitales facilita una serie de procesos:

- Calidad de los datos, a partir de la definición de un vocabulario controlado.
- La interoperabilidad a partir de herramientas de almacenamiento.
- Procesos de consultas mediante la integración de sistemas de búsquedas y procesos de exploración.
- Estructuras de conocimiento válidas y reutilizables para una variada lista de disciplinas (geografía, historia, botánica, agricultura, etc.).
- Procesos de navegación dentro de un dominio de conocimiento, a partir del uso de sus estructuras taxonómicas.

Dentro del diseño de repositorios digitales los tesauros permiten usar vocabularios controlados, con el propósito de facilitar procesos de búsqueda y clasificación de recursos digitales.

2.2.3 Ontologías

En términos generales, la ontología es una de las formas más usadas para describir formalmente el conocimiento de un dominio. Entre las definiciones más conocidas se encuentra la realizada por (Neches et al., 1991):

“una ontología define los términos básicos y relaciones que conforman el vocabulario de un área específica, así como las reglas para combinar dichos términos y las relaciones para definir extensiones de vocabularios”.

Por su parte, Gruber (1993), define una ontología como:

“una especificación explícita de una conceptualización”.

Definición base para el desarrollo de varias propuestas en el área de la inteligencia artificial, centradas en caracterizar la ontología como un instrumento que define una serie de términos básicos y relaciones a partir de un vocabulario en un área de conocimiento. A partir de esta perspectiva, se derivan una serie de reglas de vinculación entre términos y relaciones para definir extensiones a un vocabulario (Corcho, Fernández, & Gómez, 2003). Por su lado (P. Castells, 2002) considera una ontología como:

“una taxonomía de conceptos con atributos y relaciones, que proporciona un vocabulario consensuado para definir redes semánticas de unidades de información interrelacionadas”.

En la literatura se puede encontrar diferentes tipos de ontologías, de acuerdo a una serie de criterios básicos de clasificación. Algunas posturas indican que estas varían de acuerdo a su forma, estructura y uso (Valente, 2005). Las orientadas al desarrollo de aplicaciones en el área de e-learning (Stojanovic, Staab, & Studer, 2001), definen tres dimensiones ontológicas asociadas con: i) el contenido (en qué consiste un material de aprendizaje); ii) su contenido (de qué forma se presenta un tema); y iii) la estructura (define un conjunto de material de aprendizaje en un curso). A partir de este último enfoque, a continuación se presentan las clasificaciones de ontologías más representativas asociadas al diseño de repositorios digitales:

- *Ontologías para representación de conocimiento*: Permiten relacionar conceptualizaciones asociadas a formalismos de representación de conocimiento (R. Davis, Shrobe, & Szolovits, 1993; Gruber, 1993; Guarino & Poli, 1995).
- *Ontologías terminológicas, lingüísticas*: Especifica a un conjunto de términos usados para representar conocimiento de un dominio específico, permitiendo clasificar los términos en varios niveles (Lindberg, Humphreys, & McCray, 1993).
- *Ontologías de dominio*: Contiene todos los conceptos asociados a un dominio particular, a partir de una estructura detallada de un contexto y subdominios asociados (Mizoguchi, Vanwelkenhuysen, & Ikeda, 1995).

El uso de ontologías en recursos digitales ayudan a aumentar la coherencia y la interoperabilidad a partir de la descripción de sus metadatos (Jovanović, Gašević, Knight, & Richards, 2007). Sin embargo, existen estudios donde indican que el uso de ontologías presenta inconvenientes para el acceso a recursos digitales (Gašević & Hatala, 2006; Li, Gasevic, Nesbit, & Richards, 2005), dado que los usuarios prefieren utilizar terminologías comunes, que familiarizarse con el uso de una estructura ontológica. Estos inconvenientes, pueden resultar un proceso bastante complejo para llevar a cabo actividades de búsquedas de recursos digitales. Una solución a esto, es el uso de ontologías especializadas, tales como la aplicación de un Sistema de Organización Simple de Conocimiento (SKOS)(Miles, Matthews, Wilson, & Brickley, 2005), para la definición de taxonomías de un dominio específico basado en un conjunto enriquecido de propiedades para definir jerarquías de conceptos (Miles, et al., 2005).

La Figura 3 presenta a modo resumen, una adaptación de una propuesta realizada por Lei (2008), sobre el soporte que presentan algunos esquemas de representación de conocimiento en cuanto a su estructura y organización de contenidos.

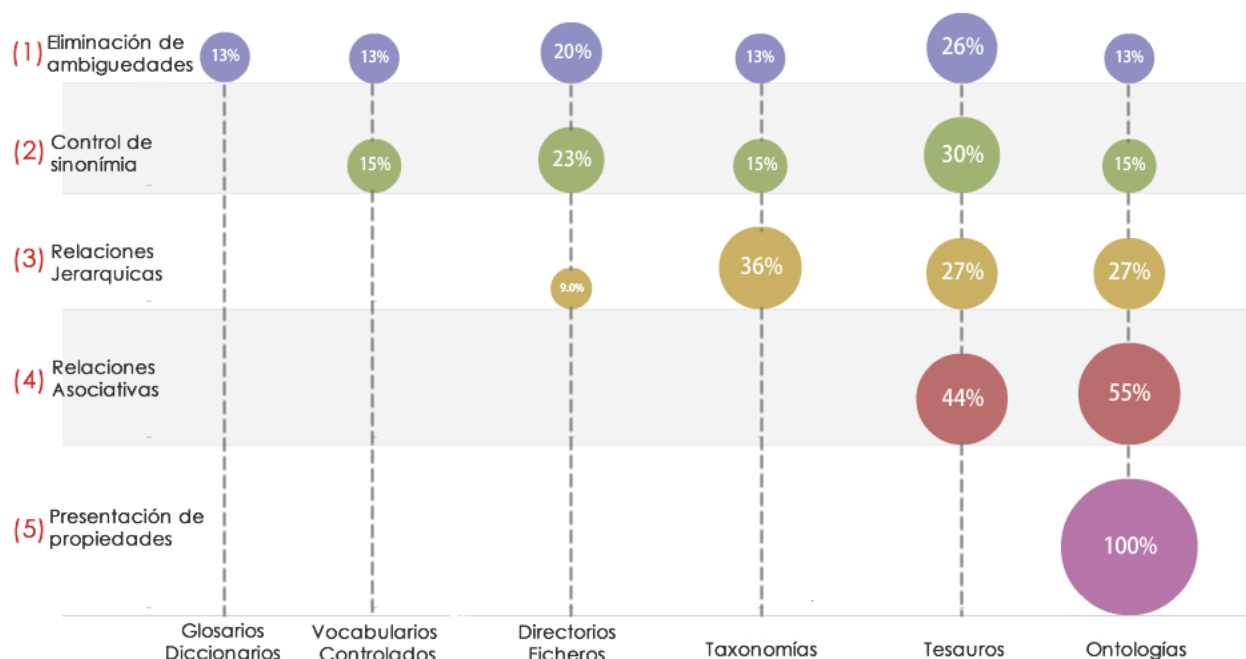








Figura 3. Estructuras y estrategias para organización del conocimiento (Lei, 2008)

En la Figura 3 se identifica en la primera columna las diferentes funcionalidades que pueden ofrecer algunos esquemas de representación de conocimiento, dentro de las cuales el autor resalta cinco principios: 1) *eliminación de ambigüedades*, asociado con la manera en que el esquema de representación facilita la ubicación de un término o concepto sin duplicaciones del término; 2) *Control de sinonimia*, se refiere al control que realiza la estructura para distinguir y/o asociar varios términos relacionados conceptualmente; 3) *Relaciones jerárquicas*, la capacidad de la estructura para determinar el nivel de relación de dos términos en función de su significado; 4) *Relaciones asociativas*, la facilidad de una estructura para indicar la relación de términos próximos entre sí para representar conceptos asociados por ideas; y 5) *presentación de propiedades*, relacionados a la facilidad de un instrumento de representación de conocimiento para asociar niveles de representación semántica a partir del uso de metadatos.

En el esquema propuesto por Lei (2008), se identifica una mayor participación de las *taxonomías*, *tesauros* y *ontologías* como estructuras que soportan mayores funcionalidades para la organización y gestión del conocimiento. A partir de estos resultados, en la Tabla 1 se presentan algunas de las funcionalidades que proporcionan estos esquemas de representación de conocimiento (taxonomías, tesauros y ontologías), para llevar a cabo actividades de organización y gestión de contenidos adaptado a partir de un estudio realizado por Soler & Gil (2010).

Tabla 1. Diferencias y funcionalidades de esquemas de representación de conocimiento (Soler & Gil, 2010)

	Taxonomías	Tesauros	Ontologías
Objetivo	Organización del conocimiento. Clasificar información	Organización del conocimiento. Indizar y recuperar información	Organización del conocimiento. Sistematizar y explorar el conocimiento
Lenguaje	Terminología comprensible por el usuario	Terminología consensuada, normalizada y controlada	Lenguaje natural, lenguaje controlado y lenguaje formal
Estructura	Jerárquica	Sistemática, jerárquica, alfabética, índice	Taxonomía, tablas con conceptos, sinónimos, descriptores, instancias, relaciones, atributos, valores, axiomas
Actualización	Inmediata y arbitraria por los diseñadores de los contenidos webs	Periódica y consensuada según la evolución de la terminología	Inmediata según la evolución del contenido de la propia ontología
Tipos de relaciones	Jerárquicas	Jerárquicas, asociativas y de equivalencia	Jerárquicas, asociativas, de equivalencia y cualquier otro tipo (temporales, causa-efecto, síntomas-tratamientos, etc.)
Aplicaciones	Presentación y organización de contenidos digitales	Indización, navegación y búsqueda	Organización, clasificación, navegación, búsqueda, interoperabilidad y razonamiento automático
Presentación	Jerárquica	Alfabética, Jerárquica, sistemática	Recomendaciones de la W3C
Inclusión de definiciones	No	Sí	Sí
Coste de mantenimiento y actualización			
Coste de elaboración			

A partir de las diferencias y funcionalidades de los esquemas de representación de conocimiento analizados por Soler & Gil (2010) (Tabla 1), y las aproximaciones planteados por Lei (2008) (Figura 3) sobre los esquemas de representación de conocimiento, finalmente en la Tabla 2 se presentan algunas de las funcionalidades más representativas que proporcionan cada uno de estos esquemas de representación de conocimiento, enfocados al desarrollo de repositorios digitales.

Tabla 2. Funcionalidades de instrumentos de representación de conocimiento

Funcionalidades	Taxonomías	Tesauros	Ontologías
Métodos de uso	Navegación	Textual y navegacional	Textual y Navegacional
Forma de Navegación	Secuencial	Secuencial - relacional	No Secuencial - relacional
Modos de acceso	Gráfico	Textual y gráfico	Textual y gráfico
Formas de búsqueda	Jerárquicas	Jerárquicas y relacional	Semánticas
Representación de Datos	Estructurada	Jerárquica, alfabética, índice	Taxonomía, tablas con conceptos, sinónimos, descriptores, instancias, relaciones, atributos, valores, axiomas
Tipos de Relaciones	Jerárquicas	Jerárquicas, asociación y equivalencia	Jerárquicas y cualquier tipo de relación
Complejidad de uso	Bajo	Medio	Alto
Nivel de interoperabilidad	Estructural	Semi estructural	Semánticas

2.3 Estrategias de acceso sobre recursos digitales

En los últimos años, los repositorios digitales han ganado un impacto representativo debido al aumento exponencial en el número de recursos digitales publicados (Margaryan & Littlejohn, 2008). Este crecimiento de recursos digitales ha generado una serie de soluciones orientadas al desarrollo de estrategias en varios frentes, a saber: i) *tecnológico*, mediante el desarrollo de repositorios distribuidos, repositorios heterogéneos y federaciones de repositorios como puntos centrales de acceso a cada uno de ellos (McGreal, 2008); ii) *semántico*, a través del uso y vinculación de sistemas de clasificación de conocimientos para proporcionar una mejor comprensión y organización de los recursos digitales (Soto, Gordo, & Sánchez, 2007); y finalmente iii) *acceso*, mediante el desarrollo de mecanismos y estrategias de búsqueda para ofrecer acceso a recursos digitales.

Uno de los desafíos que presentan actualmente los repositorios digitales no se limita solo a la gestión y administración de recursos digitales, o metadatos, según sea el caso; sino a ofrecer mecanismos de búsqueda que faciliten el acceso a cientos y miles de recursos digitales relevantes, alojados en cada uno de ellos. Por lo tanto, para llevar a cabo esta actividad, se deben analizar, identificar y evaluar una serie de factores que permitan mitigar estas limitaciones. En la Figura 4, se realiza una adaptación de los factores más representativos, de acuerdo a criterios definidos en estudios previos para el acceso a una colección de recursos

digitales (Card, Mackinlay, & Shneiderman, 1999; Marchionini, 2006, 2008; Petrelli, 2008; Shneiderman, 2003).

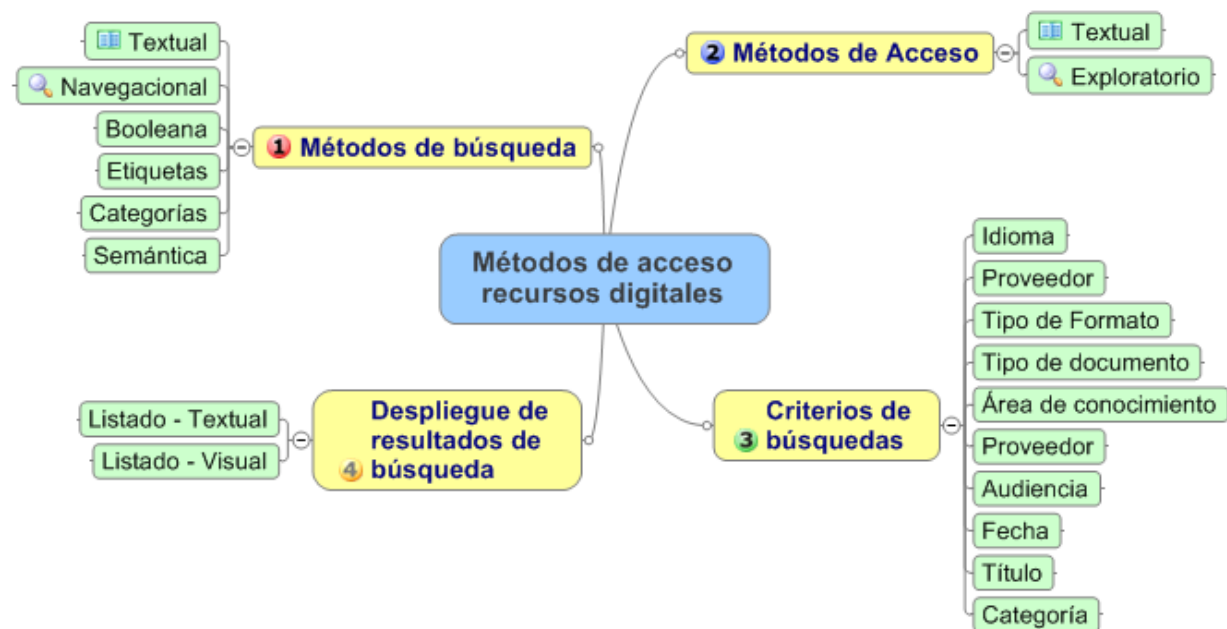


Figura 4. Métodos de acceso sobre recursos digitales

En los siguientes apartados se comentan cada uno de estos factores de acceso más relevantes.

2.3.1 Métodos de búsqueda

En general, los sistemas de búsqueda de repositorios digitales proporcionan al menos dos alternativas de búsqueda a partir de su interfaz de usuario: *búsqueda básica o simple* y *búsqueda avanzada*. La *búsqueda básica* no requiere que el usuario esté bien informado sobre el sistema y el proceso de búsqueda. Permite al usuario realizar una búsqueda rápida, pero restringe al usuario mediante el uso palabras claves. Se supone que es fácil de entender y de utilizar por parte de usuarios sin una previa experiencia de uso. Por su parte, la *búsqueda avanzada*, requiere más conocimientos y habilidades de búsqueda por parte del usuario.

Las búsquedas básicas son las más populares, mientras que las búsquedas avanzadas carecen de facilidades de comprensión para su uso por parte de un usuario, dada la complejidad que presentan a partir de las características definidas en algunas de ellas. Sin embargo, el diseño de interfaces juega un papel importante en este proceso, ya que facilitaría en gran medida actividades de localización de recursos digitales, y mejoraría notablemente la satisfacción de un usuario (Shneiderman, 2003). Dentro de los métodos más convencionales definidos para realizar procesos de búsqueda sobre recursos digitales en repositorios, se pueden destacar las siguientes:

- *Búsqueda textual (palabras claves)*. Normalmente se encuentra asociada a un conjunto de palabras claves, vinculadas a través de un conjunto de metadatos definidos en cada recurso digital.
- *Búsqueda navegacional (exploración)*. Es un tipo de búsqueda exploratoria, basado en una serie de categorías y clasificaciones definidas por una estructura jerárquica por niveles. Por lo general, la forma de despliegue se realiza mediante carpetas o árboles. Cada categoría permite desplegar una serie de recursos asociados, a la medida que el usuario realiza el proceso de exploración.
- *Búsqueda booleana (lógicas)*. Definido a partir de filtros de búsqueda lógicas de tipo (AND/OR), mediante la asociación de palabras claves y algunos criterios de selección.
- *Búsqueda etiquetada*. Basado a partir de la definición de nubes de palabras claves más usadas por usuarios para llevar a cabo un proceso de búsqueda.
- *Búsqueda por categorías*. Se define a partir de temáticas o secciones que permiten categorizar un área de conocimiento.
- *Búsqueda semántica (asociación y clasificación)*. Es un tipo de búsqueda basado en un lenguaje semántico definido a través de un esquema de representación de conocimiento. Una búsqueda semántica denota un concepto sobre el que un usuario está tratando de obtener información (Guha, McCool, & Miller, 2003; Mangold, 2007). De esta manera, es posible desarrollar herramientas de búsqueda que entienden la semántica de una consulta con el fin de responder a ciertas necesidades. Para llevar a cabo este proceso, se parte del uso de esquemas de representación de conocimiento, definidos a partir de instrumentos como las ontologías y los tesauros. Las búsquedas semánticas mantienen una serie de relaciones entre recursos digitales que disponen de un nivel de asociación. Esto permite vincular únicamente aquellos recursos digitales que tengan un nivel de afinidad, de acuerdo a la consulta definida por el usuario. Para facilitar su uso, es necesario contar con la definición de una estructura taxonómica que permita jerarquizar y determinar los niveles de profundidad del área de conocimiento explorado.

2.3.2 Métodos de acceso

De acuerdo a los propósitos de esta investigación, se asocia el concepto de accesibilidad a la facilidad para ingresar a recursos digitales en un repositorio, desde el punto de vista de la usabilidad de las interfaces (ver sección principios de usabilidad Sección 2.4.3), y no conforme a las estrategias o métodos definidos para el acceso de personas discapacitadas definidos según la UNE 139803:2004. Por lo tanto, dentro de los métodos más comunes definidos para realizar procesos de acceso a recursos digitales en repositorios podemos destacar los siguientes:

- *Método tradicional (búsqueda textual)*. Definido a partir de palabras claves y despliegue de resultados mediante un listado de recursos digitales. La Figura 5, presenta un ejemplo de este método de búsqueda.

The screenshot displays the VOA3R repository's search interface. On the left is a navigation menu with links: Home, Social search, Find Resources (highlighted), Navigational Interface, By Author, Browse, Timeline view, Map view, Partners, Become a partner!, and About us. The main search area is titled 'Text based search' and includes a 'Search Keyword' input field with the placeholder 'Keyword' and a 'Search' button. Below this is the 'Advanced search' section, which contains a 'Search for:' label, three radio buttons for 'Natural search' (selected), 'All words', and 'Exact Phrase', and two dropdown menus for 'Ordering field' (set to 'relevance') and 'Ordering' (set to 'Descendent'). A checkbox labeled 'Show FULLTEXT result first' is checked. At the bottom is the 'Filters' section with dropdowns for 'By repository' (set to 'Any') and 'By language' (set to 'Any'), and a date range filter 'By date: from' to 'to' with empty input fields. The text 'Búsqueda textual' is written at the bottom of the interface.

Figura 5. Ejemplo de métodos de búsqueda textual repositorio VOA3R

La Figura 5 presenta un ejemplo de búsqueda textual combinada con un campo de búsqueda básica y un conjunto de opciones de búsqueda avanzada, a partir de ciertos criterios definidos en el repositorio digital VOA3R.

- *Método exploratorio (esquema de navegación visual)*. Permite realizar actividades de navegación a través de una estructura jerárquica o taxonómica que permite visualizar los niveles y clasificaciones. Ofrece la posibilidad de llevar a cabo un proceso de aprendizaje y finalmente seleccionar las categorías de interés para el despliegue de resultados. La Figura 6, presenta a modo de ejemplo este método de búsqueda.

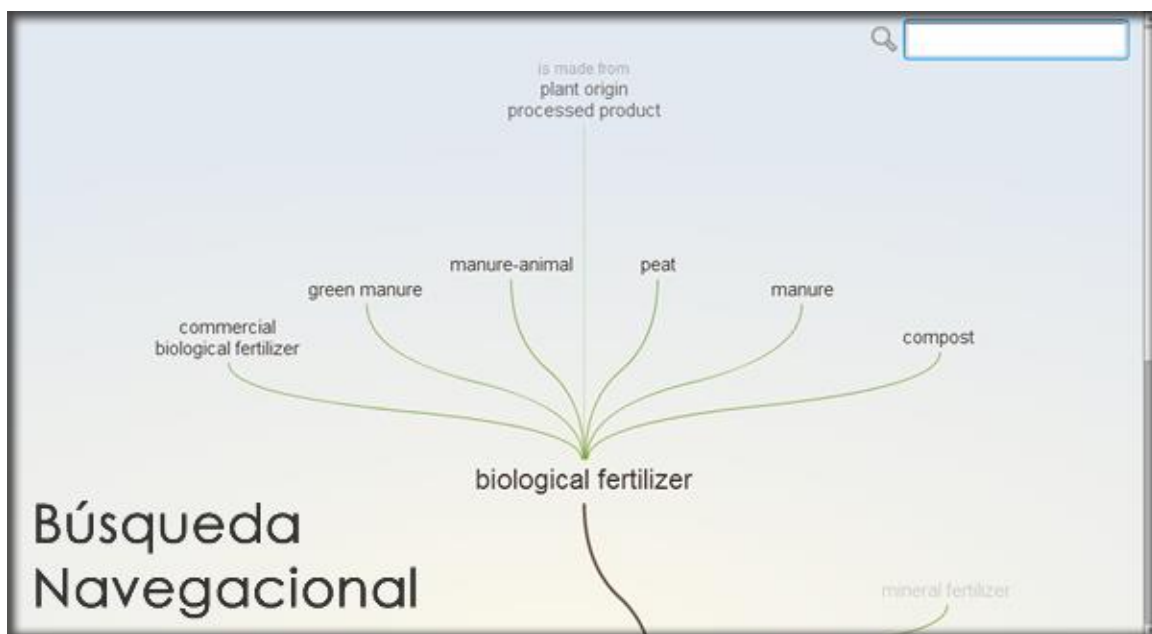


Figura 6. Ejemplo Métodos de búsqueda navegacional repositorio Organic.Edunet

La Figura 6, es una interfaz de búsqueda navegacional, a partir de la estructura de un árbol jerárquico definido en el repositorio digital Organic.Edunet, donde representa una jerarquía taxonómica de temáticas asociadas a fertilizantes biológicos.

2.3.3 Criterios de búsqueda

Los criterios de búsqueda permiten definir y parametrizar las consultas de acuerdo a las necesidades del usuario. Estas varían de acuerdo a la definición del recurso digital y las descripciones asociadas a sus metadatos. Por ejemplo: área de conocimiento, idioma, proveedor de contenido, formato del recurso digital (imagen, video, audio, etc.). En la Figura 7 se presenta un ejemplo de varios criterios de búsqueda definidos para una colección de recursos digitales sobre la biblioteca digital Europea.

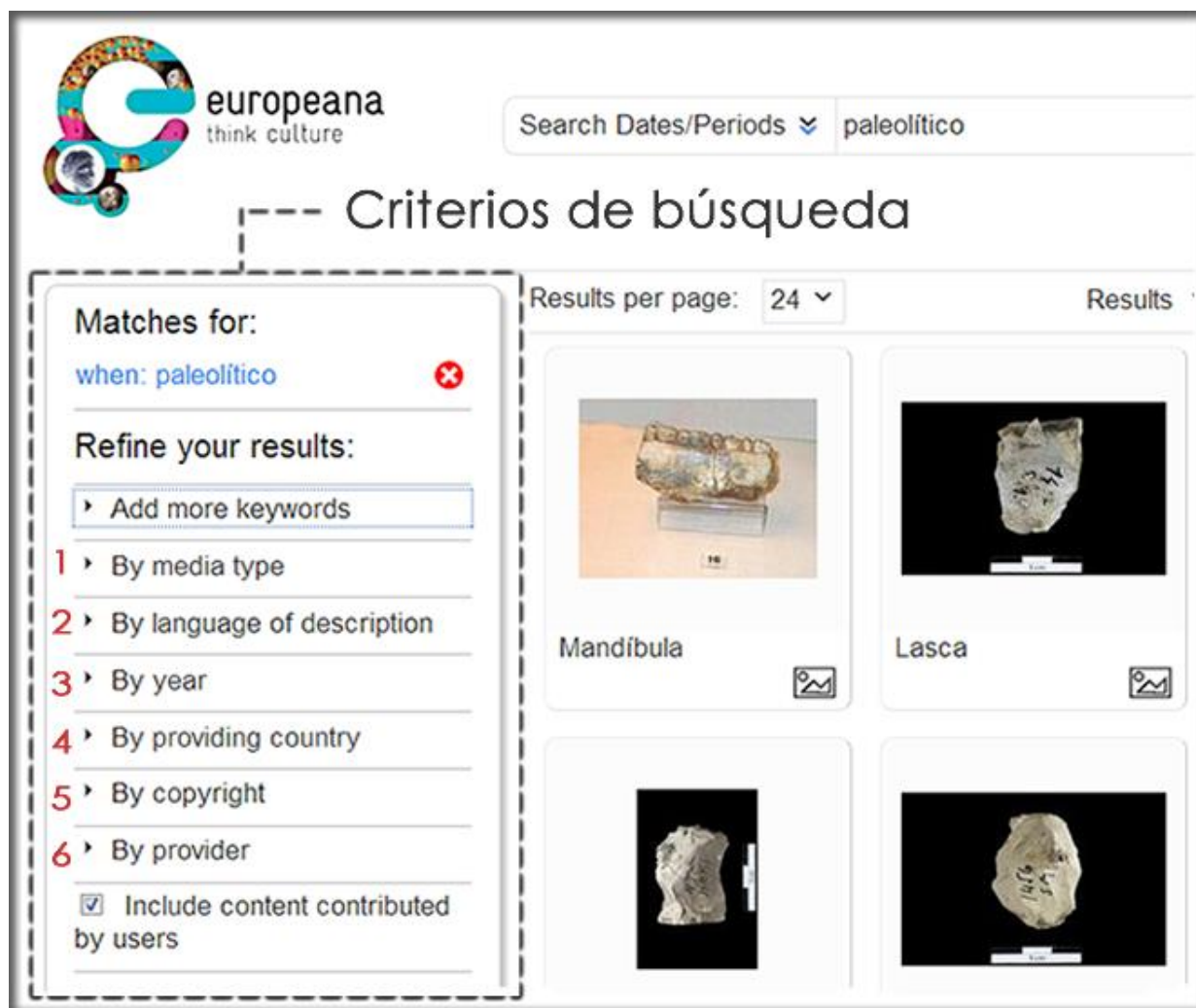


Figura 7. Ejemplo de criterios de búsqueda definidos en biblioteca digital Europeana

Para el caso de la Figura 7, se presentan una serie de criterios de búsqueda de recursos digitales a partir de la definición de metadatos. Para este caso se identifican atributos asociados con: 1) tipo de formato digital, 2) lenguaje, 3) año, 4) país, 5) derechos de autor y 6) proveedor de contenidos. Este factor se complementa con mayor detalle en la Sección 2.4.3 y a través de las estrategias usadas para el estudio, definidos en el Capítulo 4.

2.3.4 Despliegue de resultados de búsqueda

Es la manera en que la interfaz despliega los resultados de búsqueda de las consultas realizadas por el usuario. Estas estrategias por lo general están asociadas con la carga de un listado de recursos digitales mediante métodos de paginación. Suelen desplegarse de manera textual, o gráfica, a partir de la pre-visualización de resultados asociados a una imagen e información complementaria.

Para evaluar el uso y satisfacción de este tipo de estrategias a la hora de mostrar resultados, se emplean criterios de usabilidad. La Figura 8, presenta un ejemplo de la estrategias de despliegue en lista de manera textual usado en el repositorio digital MERLOT.

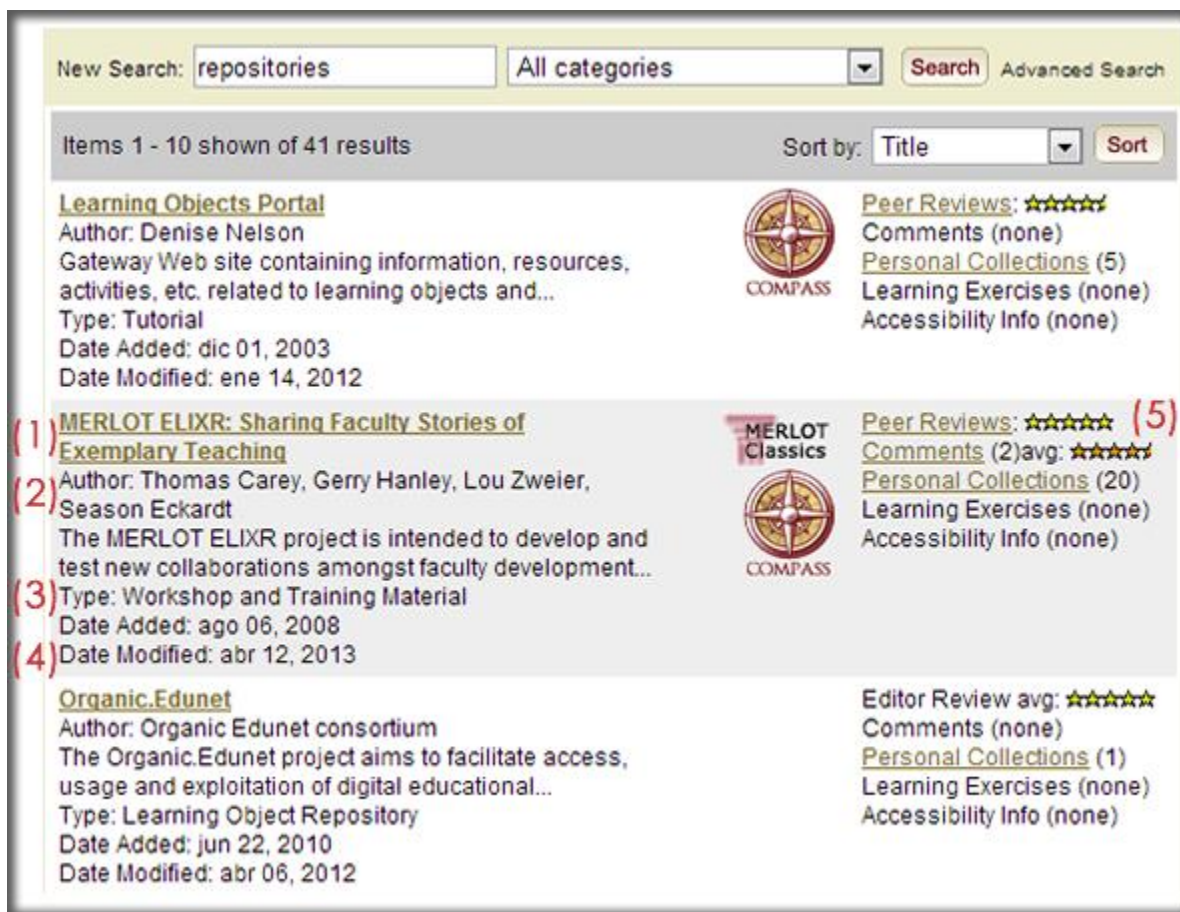


Figura 8. Ejemplo del despliegue de recursos digitales tipo lista - textual usado en el repositorio digital MERLOT

Este tipo de despliegue, tiene como característica mostrar información relevante del recurso digital, previo al acceso del mismo. Para el caso del repositorio digital de MERLOT, como se puede apreciar en la Figura 8, se presenta un despliegue de información relacionada con: 1) el título, 2) autor, 3) tipo de recurso digital, 4) fecha de actualización y 5) valoración del recurso digital.

Otro tipo de despliegue de información de un recurso digital se puede ver en la Figura 9, mediante una combinación de un listado de recursos digitales textual y gráfico.

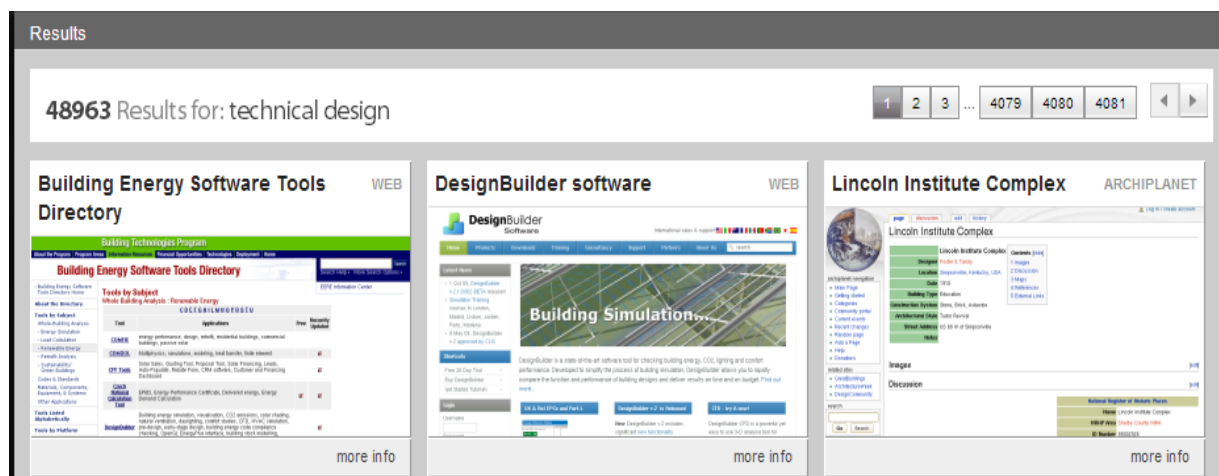


Figura 9. Ejemplo despliegue de recursos digitales tipo lista, pre-visualización del repositorio MACE

La Figura 9 presenta a modo de ejemplo, un despliegue de recursos digitales del repositorio digital MACE, donde el listado de recursos digitales se combina con la pre-visualización del recurso digital.

Finalmente en la Tabla 3 se presenta una clasificación sobre las estrategias de acceso implementadas por algunos de los repositorios digitales más relevantes.

Tabla 3. Estrategias usadas por repositorios digitales para acceder a recursos digitales

Repositorio	Método de búsqueda	Método de Acceso	Criterios de búsqueda	Estrategia de despliegue consultas
E-LERA^[1]	Textual	Textual	Área de conocimiento	Listado - textual
MERLOT^[2]	Categorías Textual	Textual	Área de conocimiento	Listado - textual
ARIADNE^[3]	Textual	Textual	Proveedor Idioma Tipo de Formato Tipo de documento	Listado - textual
MACE^[4]	Textual Categorías Semántico Etiquetas	Textual Árbol Facetas	Área de conocimiento Ubicación Competencia Comunidad social	Listado - textual y gráfico

Repositorio	Método de búsqueda	Método de Acceso	Criterios de búsqueda	Estrategia de despliegue consultas
Organic.Edunet ^[5]	Textual	Textual	Contexto educativo Idioma	Listado - textual
	Semántico	Árbol	Tipo de Formato Audiencia	
E-prints ^[6]	Textual	Textual	Materias Año Autor Tipo de documento Facultad, departamento	Listado - textual
			Autor Materias Fecha Comunidades	
Dspace ^[7]	Textual	Textual	Tipo de Formato Tipo de documento	Listado - textual
CERN ^[8]	Textual Categorías	Textual	Comunidades Colecciones Fecha Autores Títulos Temas	Listado - textual
MIT ^[9]	Textual	Textual	Área de conocimiento Fecha de publicación Tipo de documento Departamento Tipo de Formato Categoría Colecciones	Listado - textual
NASA ^[10]	Textual Categorías	Textual		Listado - textual

Como se ha comentado, las interfaces visuales son una buena alternativa para facilitar las búsquedas y por ello en el siguiente apartado se llevará a cabo una mayor profundización, analizando la influencia de las técnicas de visualización de información, tanto para el despliegue de resultados como para acceder a ellos.

2.4 Visualización de información en repositorios y bibliotecas digitales

Para llevar a cabo procesos de búsqueda de objetos de aprendizaje, los repositorios a menudo proporcionan alternativas de acceso, algunas de ellas "visuales" mediante el uso de interfaces. En esta sección se resalta la importancia que presenta el área de visualización de información en procesos de búsqueda y acceso a recursos digitales. Para ello, se presenta en primera instancia la importancia de la visualización de información en el ámbito informático, seguido del proceso que se lleva a cabo para implementar este tipo de estrategias y las áreas involucradas dentro del proceso de evaluación. Finalmente, se mencionan los primeros trabajos sobre visualización de información, relacionados en el campo de repositorios y bibliotecas digitales.

2.4.1 Visualización de Información

La visualización de información es la representación visual de información compleja haciendo uso adecuado de espacios y estructuras gráficas (Herman, et al., 2000). Su propósito es facilitar una rápida asimilación y comprensión de la información. Una de las definiciones clásicas sobre visualización de información en la literatura fue realizada por Card (1999) como:

“el uso de representaciones visuales de datos abstractos, interactivas y basadas en el uso del ordenador para amplificar el proceso cognitivo”.

La visualización de información tiene varias áreas de estudio, los cuales han sido abordados desde la percepción gráfica del usuario para determinar las influencias asociadas a variables visuales tales como la posición, longitud, color y forma, que influyen en la eficacia de la visualización de datos (Cleveland, 1984; Simkin & Hastie, 1987; Vessey & Galletta, 1991).

La visualización de información presenta una variada lista de beneficios para lograr representar información compleja. De acuerdo con (González & Kobsa, 2003), en el área de gestión y administración identificamos algunas de las más relevantes:

- *La flexibilidad en el manejo de datos.* Mediante técnicas de zoom, las personas tienen una mayor flexibilidad en el manejo de información. Esto les permite centrarse en información relevante en ciertas áreas dentro de la representación gráfica desplegada.
- *Imagen completa de los datos completos.* Facilita una idea clara de los datos en todo su contexto para una fácil interpretación de los mismos.
- *Fácil extracción de gráficos e informes.* Ofrece una variada lista de posibilidades para facilitar la extracción de informes.

La visualización de información ha marcado una fuerte línea de trabajo, estableciendo disciplinas de mayores trayectorias en el ámbito de la computación gráfica. Una de ellas se conoce como Interacción Persona Ordenador (IPO). Esta disciplina, conocida globalmente

como Human Computer Interaction (HCI), desde sus orígenes ha permitido centrar la atención en el desarrollo de productos informáticos, atendiendo a necesidades básicas de uso y tiempos de ejecución. De todo ello se llevará a cabo una breve introducción en el siguiente apartado.

2.4.2 Interacción persona ordenador

Dentro de la toma de decisiones, se han argumentado que las técnicas de visualización de información deben diseñarse de acuerdo con los principios cognitivos y de percepción (Lee & Vickers, 1998; Wise, 1999). En la disciplina de análisis y evaluación de interfaces computacionales, muchas de las evaluaciones de técnicas de visualización han descuidado las capacidades de un usuario para utilizar una técnica de manera precisa y eficiente (Tractinsky, Katz, & Ikar, 2000). Una de las estrategias para evaluar estas interfaces es mediante la interacción de personas de forma empírica, es decir, probando la manera en que las personas utilizan estas interfaces de manera controlada y experimental (Allan, Leuski, Swan, & Byrd, 2001; Lee, Reilly, & Butavicius, 2003; Petrelli, 2008; Sebrechts, Cugini, Laskowski, Vasilakis, & Miller, 1999; Sutcliffe, Ennis, & Hu, 2000; R. Swan & Allan, 1998; Westerman & Cribbin, 2000). Esta experimentación es crucial debido a la evidencia de que las evaluaciones objetivas y subjetivas a menudo difieren en las evaluaciones de interfaz (Frøkjær, Hertzum, & Hornbæk, 2000; Lee, et al., 2003; Wu, Fuller, & Wilkinson, 2001).

La Interacción Persona Ordenador (IPO) es una disciplina en constante crecimiento, que desde sus orígenes ha permitido abrir un espacio en el desarrollo de productos informáticos, cuyo objetivo es atender las necesidades de uso y tiempos de ejecución, para evitar así posibles errores y en consecuencia ofrecer satisfacción del producto a los usuarios (Booth, 1989; Dix, Janet, Abowd, & Beale, 2003; Marchionini, 1997; Shneiderman, 2003). La visualización de información presenta varias áreas de estudio, los cuáles se han abordado desde el campo de la computación gráfica y de la interacción persona ordenador, mediante investigaciones orientadas a determinar las influencias asociadas a través de variables visuales como la posición, longitud, color y figura, que impacten en la efectividad de una visualización de datos (Cleveland, 1984; Simkin & Hastie, 1987; Vessey & Galletta, 1991). Otros estudios más recientes en este campo se involucran con técnicas avanzadas de análisis visual como *eye-tracking* (Jacob & Karn, 2003; Poole & Ball, 2006; Rosch & Vogel-Walcutt, 2013), estrategias para evaluación de diseños de productos en la Web (J. Heer & Bostock, 2010; Simkin & Hastie, 1987), y una variada lista de aplicaciones que se encuentran registradas en uno de los más recientes análisis de investigación orientados en esta línea (von Landesberger et al., 2011). Para llevar a cabo este tipo de estudios, una de las áreas encargadas de analizar este tipo de comportamientos se conoce como la usabilidad.

2.4.3 Principios de usabilidad

La usabilidad es un concepto que se refiere principalmente a la facilidad de uso de una aplicación o producto interactivo. Hay una variedad de estrategias para evaluar este campo de estudio. Empíricamente, la usabilidad puede ser medida y evaluada. Por lo tanto, la usabilidad

es un atributo de calidad que se acompaña de diversos componentes (Nielsen, 1994b). Algunos de los más importantes son:

- *Eficacia*: durante la realización de una actividad, determinar el número de errores que hace un usuario, o de la confianza para llegar a realizar una tarea.
- *Eficiencia*: define el tiempo que lleva a los usuarios realizar alguna actividad.
- *Satisfacción*: se determina cómo de agradables y simples se perciben las actividades de desarrollo de un usuario.
- *Facilidad de uso*: medida del nivel de utilidad que puede tener la herramienta para llevar a cabo una actividad.

Existen principios básicos para el diseño de interfaces de modo que se facilite su usabilidad. Uno de ellos es la *simplicidad*, la cual se puede definir como la facilidad que ofrece una interfaz para realizar alguna actividad en particular, sea de búsqueda, exploración, etc. Nielsen (2006) se refiere al concepto de simplicidad como:

“una práctica continuada, centrada en optimizar los sitios Web y que permita maximizar el éxito de los negocios”.

Práctica, que se considera como una forma de reducir las barreras que se imponen a los usuarios para llegar a donde deseen, dentro de una aplicación informática.

En el proceso de usabilidad, se persigue también la reducción del número de procesos que debe realizar un usuario para llevar a cabo determinada actividad; pues tal como lo reseña (Norman, 2008), el objetivo final de la usabilidad es reducir la frustración del usuario a la hora de enfrentarse a las tareas. Por lo tanto, la simplicidad también desempeña un papel importante para lograr un diseño con buen nivel de comprensión, y por tanto, que permita generar experiencias de uso satisfactorias. Hearst (2009) definió la importancia de los pequeños detalles de diseño de la interfaz, tales como el tamaño del cuadro de búsqueda, la existencia de sugerencias de ortografía, etc., detalles que pueden afectar profundamente la conducta de búsqueda de información por parte de los usuarios.

Además de estas directrices para medir usabilidad, existen propuestas concretas de principios de usabilidad centrados en niveles de consistencia (Shneiderman, 2003), factores heurísticos (Nielsen, 1994b), criterios para identificación de errores y métodos de recuperación de problemas de uso (Polson & Lewis, 1990) o niveles de flexibilidad (9241 ISO, 1998), entre otros.

Vistos los conceptos básicos sobre visualización de información y los principios a tener en cuenta, a continuación se centra la revisión en la visualización dentro del ámbito de las bibliotecas digitales.

2.4.4 Visualización de información en bibliotecas digitales

Uno de los primeros intentos de desarrollar interfaces mediante estrategias de visualización para vincular recursos digitales dentro del campo de la recuperación de documentos, se llevó a cabo a través del Science Citation Index (SCI). Esta herramienta permitió la vinculación de los índices de citas de autores en revistas científicas (Mora, 2004).

Con el tiempo, el uso de técnicas de visualización se aplicó en otras áreas de la documentación. Uno de los campos de mayor trascendencia es el de sistemas de recuperación de información, también conocidos como Visual Information Retrieval Interfaces (VIRI). Bajo esta línea de trabajo, se llevaron a cabo varias propuestas de desarrollo a partir del uso de visualización. Por un lado se presentaron enfoques orientados a la visualización de colecciones de recursos digitales a través de mapas conceptuales (X. Lin, 1997; X. Lin, Soergel, & Marchionini, 1991). Otros autores se centraron en el desarrollo de estrategias de búsqueda a partir de métodos de agregación de recursos digitales (Cugini, Laskowski, & Sebrechts, 2000; Foltz, 1997); estas estrategias favorecieron la vinculación de atributos del recurso digital mediante las descripciones definidas en cada uno de ellos (Hearst, 1995; Veerasamy & Belkin, 1996). De las anteriores propuestas, se pueden indicar que fueron estrategias de visualización orientadas a facilitar procesos de despliegue de información, mediante el uso de bases de datos y catálogos de archivos.

Para acceder a un gran volumen de información, varias propuestas han dedicado importantes esfuerzos en estudiar diferentes enfoques para el diseño de interfaces visuales (Allen, Obry, & Littman, 1993). A continuación presentamos un resumen de los estudios más relevantes sobre esta línea de trabajo: VR-VIBE (Benford et al., 1995), un prototipo de colaboración para buscar recursos de una biblioteca y desplegar conceptos representados mediante efectos visuales sobre tres dimensiones; LyberWorld (Hemmje, Kunkel, & Willett, 1994), un sistema de visualización que utiliza técnicas geométricas de diagramas mediante conos y sistema de navegación a partir de una estructura definida por un árbol (Robertson, Mackinlay, & Card, 1991), lo cual permite el despliegue de rutas de navegación basados en estrategias de búsqueda de secuencia; Cat-a-Cone (Hearst & Karadi, 1997), un prototipo que integra la búsqueda y la navegación a través de una amplia colección de recursos digitales, mediante la visualización de consultas tridimensionales y representaciones jerárquicas; DARE (J. Zhang & Korfhage, 1999), es otro ejemplo de un prototipo que utiliza representaciones angulares para formar una visualización en dos dimensiones de una colección de recursos digitales y, por último TOFIR (J. Zhang, 2001), una herramienta diseñada para visualizar información a través representaciones angulares y facilitar procesos de búsqueda de documentos.

Recientemente, en el campo de la recuperación de la información, se han planteado varias alternativas para mejorar la interacción de los registros en una biblioteca digital. Estas propuestas se han centrado en estrategias de visualización para apoyar actividades de exploración y descubrimiento (Merčun & Žumer, 2010). Propuestas que se derivan de estos análisis se incluyen el proyecto FrbrVis (Merčun & Žumer, 2009, 2010; Merčun, et al., 2012), un prototipo desarrollado a través de cuatro técnicas de visualización (*árbol*, *sunburst*, *radial-Tree*

y diagramas de *árbol circular*) para acceder a materiales de una biblioteca. Por lo tanto FrbrVis, ha sido un proyecto específicamente orientado al análisis de navegación a partir de las estructuras jerárquicas, así como las relaciones definidas entre cada uno de los registros consultados.

Una vez comentadas algunas de las estrategias posibles de visualización para vincular recursos digitales, a continuación se estudian algunas técnicas particulares de visualización.

2.5 Técnicas de visualización

La información, en el mejor de los casos presenta una estructura que permite definir una clasificación a partir de una jerarquía. Trabajos previos sobre estas estructuras se han centrado en abordar problemas relacionados con el tamaño, el diseño y la visibilidad limitada que se presenta en un área de trabajo (Battista, Eades, Tamassia, & Tollis, 1998). Existe información que presenta estructuras más complejas para su representación, y que por lo tanto requiere de técnicas de visualización que permitan resaltar los aspectos más importantes. En este sentido, las estructuras de visualización de información y sus representaciones en pantalla (Card, et al., 1999), han sido uno de los enfoques de mayor crecimiento y aplicación en el área de la visualización, siendo las estructuras de navegación las más elegidas para llevar a cabo actividades de despliegue de información (Plaisant, Grosjean, & Bederson, 2002).

Existen trabajos que han permitido acercarse al estudio de las diferentes formas de representar información mediante técnicas de visualización (M. Graham, J. Kennedy, & D. Benyon, 2000; Herman, et al., 2000). En estos estudios se realiza una descripción detallada de las diferentes propuestas que se han abordado en el área de la visualización de información. En el siguiente apartado, se identifican algunas de las estrategias de mayor relevancia en el campo de la visualización, sus implicaciones y características más representativas. Esta clasificación ha sido adaptada a partir de estudios previos realizados por Graham (2010), (Draper, Livnat, & Riesenfeld, 2009) y (Gleicher et al., 2011).

2.5.1 Visualización mediante árboles

Es un tipo de visualización clásica implementada para ubicar recursos de manera jerárquica por niveles de clasificación. Se considera clásica porque es uno de los estilos de representación más comunes difundidos a partir del uso de interfaces gráficas y navegación de carpetas mediante el explorador usado por Microsoft Windows (Kobsa, 2004). En un estudio empírico, se logró comprobar la preferencia de este estilo de representación, en comparación con otros estilos de visualización de tipo árbol (Cockburn & McKenzie, 2000).

En su esencia, ha sido una de las estructuras de navegación más investigadas (Card, et al., 1999; Chen, 2004), y con un amplio uso dentro de la rama de la computación gráfica (M. Graham, et al., 2000; Herman, et al., 2000). Desde una perspectiva taxonómica, las estructuras

de navegación se utilizan para describir los vínculos explícitos entre los nodos en diferentes árboles de clasificación (Graham & Kennedy, 2005).

De acuerdo con Graham (2010), el diseño de árboles simples se divide en varias categorías, de acuerdo a su representación gráfica para indicar una relación entre padre e hijos. En la Figura 10, se muestra una representación de dos formas comunes de representación de árboles. La primera -Figura 10 (a)-, es la forma convencional de clasificación de un árbol para desplegar las relaciones entre padres e hijos. En la segunda -Figura 10 (b)-, se representa en forma de listados, desplegadas de manera vertical, para representar el nivel de profundidad y nodos del árbol.

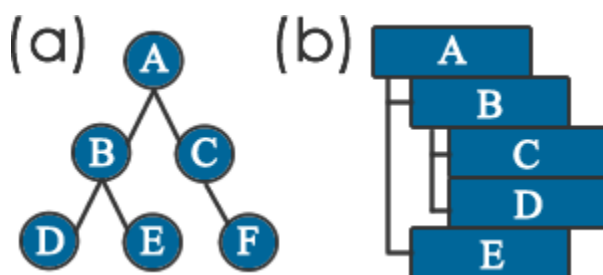


Figura 10. Formas básicas de representación de árboles (Graham & Kennedy, 2010)

Existen clasificaciones de árboles que manejan una serie de nodos únicos en sus interiores a partir de un nodo base; a estas ramificaciones se denominan multi árboles (Furnas & Zacks, 1994; Wittenburg, Das, Hill, & Stead, 1995). Este tipo de representación, es una de las más usadas para poder visualizar largas colecciones de recursos en bibliotecas digitales y repositorios digitales. En la Figura 11, se muestra una adaptación del uso de un árbol básico mediante una estructura de representación taxonómica.

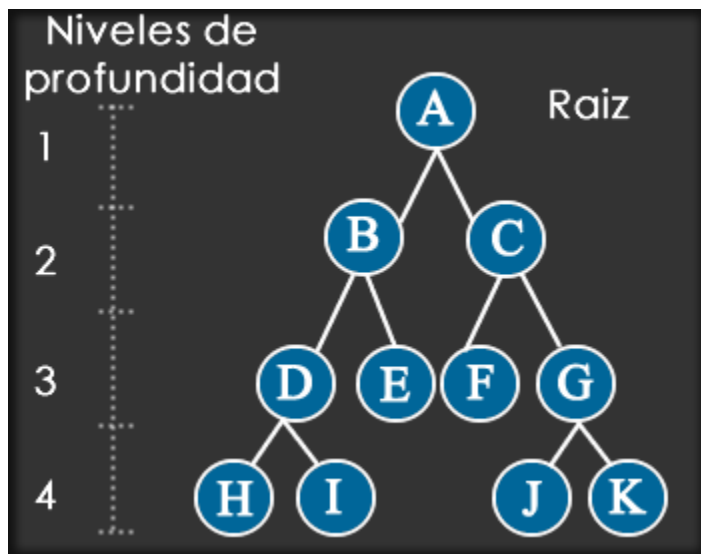


Figura 11. Adaptación de técnica de visualización tipo árbol

2.5.2 Visualización de tipo radial

El término de visualización radial aparece por primera vez a través de Hoffman en el año de 1997, mediante la exploración de técnicas para el diseño y clasificación de secuencias de ADN enmarcado en el proyecto del Genoma Humano (Hoffman, Grinstein, Marx, Grosse, & Stanley, 1997). Desde sus comienzos, ha sido una técnica ampliamente difundida en análisis estadísticos mediante gráficos circulares de *tipo pastel* (Spence, 2005) y *de tipo radial* (Wilkinson & Wills, 2005). Esto ha permitido la vinculación de componentes gráficos para la representación de nodos. Para llevar a cabo esta representación, utiliza enlaces que permiten identificar una estructura de navegación de acuerdo a una clasificación definida.

Por lo general, el posicionamiento radial presenta nodos en círculos concéntricos en función de su profundidad en la estructura taxonómica (Eades, 1990). Esto permite tener un panorama global de los niveles de clasificación que se pueden abordar. Bajo esta categoría se encuentra un listado de múltiples propuestas que despliegan información de manera radial, algunos de los ejemplos que podemos encontrar bajo esta categoría son: Visual Thesaurus (Design, 1998), 1998), Radial Traffic Analyzer (Keim, Mansmann, Schneidewind, & Schreck, 2006) y VisAlert (Livnat, Agutter, Moon, Erbacher, & Foresti, 2005). La Figura 12 presenta un acercamiento de este tipo de visualización en su esencia.



Figura 12. Adaptación de técnica de visualización radial

En esta Figura 12 se muestra una adaptación de este tipo de técnicas radiales, identificando la raíz en la parte central de la estructura de navegación, y de manera radial se despliegan las demás categorías hacia el exterior. Uno de los trabajos más representativos que permiten identificar un panorama más amplio del uso de este tipo de técnicas, es el estudio realizado por Draper (2009), donde se realiza una revisión sobre la evolución que ha tenido este tipo de visualizaciones y se relaciona los problemas más frecuentes junto con una serie de propuestas

de diseños basados en patrones a partir del uso de jerarquías taxonómicas para tratar de contrarrestar este tipo de problemas.

2.5.3 Visualización mediante interfaces hiperbólicas

Son estructuras de tipo circular, que trabajan a partir del manejo de técnicas de foco y contexto, basados en la geometría hiperbólica para la visualización y manipulación de grandes jerarquías (L. Lamping & Rao, 1996). Aun siendo una derivación de la visualización de tipo radial, se ha definido en otra categoría por su estructura de navegación y el despliegue de nodos y discos.

Dentro de esta categoría se encuentran propuestas de visualización conocida como *sunburst* (Stasko, Catrambone, Guzdial, & McDonald, 2000). Una técnica que utiliza un radio en lugar de un diseño rectangular. En *sunburst* los artículos de una jerarquía están organizados de manera radial, la parte superior de la jerarquía se presenta en el centro y los niveles de clasificación más profundos se despliegan lo más lejos del centro. Sunburst trabaja con técnicas de foco y contexto y con estrategias de interacción mediante *zooming* y *panning* de manera elíptica. Al trabajar radios concéntricos permite realizar distribución de directorios de manera jerárquica. La profundidad se representa por niveles y se presentan de manera secuencial lo más lejano del centro. Permite la visualización total de la estructura de navegación. En la Figura 13 se presenta un ejemplo.



Figura 13. Visualización sunburst

La Figura 13 presenta una clasificación de electrodomésticos clasificados por función dentro de un hogar. En la parte central se identifica el nivel principal “hogar”, el siguiente radio presenta el nivel de clasificación en función de la funcionalidad del electrodoméstico dentro del hogar; y finalmente los niveles 3 y 4 presentan el nombre del electrodoméstico.

Dentro de esta categoría también se han definido técnicas de visualización basadas en la selección de nodos mediante técnicas de *zooming* y *panning* (hace referencia a un efecto que permite enfocar un objeto relevante, y desenfocar objetos externos). Algunas de ellas utilizan técnicas visuales de *foco* y *contexto* (Pirolli, Card, & Van Der Wege, 2001) representada a través del efecto ojo de pez (Furnas, 1986; Noik, 1993). Un ejemplo de ello, es la técnica conocida como *hypertree*; estrategia que imita el efecto que tiene una cámara fotográfica al ampliar la vista del nodo que se está consultando y ocultando el resto de las categorías representados en su estructura de navegación. Permite la visualización de términos hasta tres niveles de profundidad. Una adaptación de este tipo de técnicas se muestra en la Figura 14.

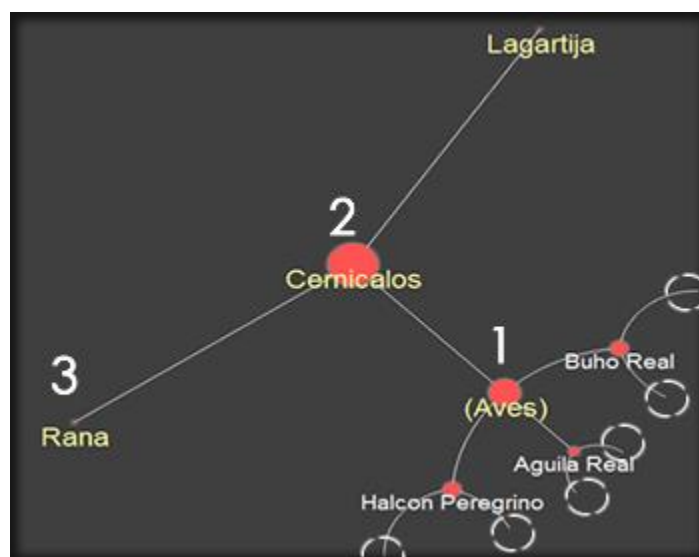


Figura 14. Visualización hypertree

La Figura 14, presenta un ejemplo del uso de esta interfaz mediante una representación de aves de presa. Los números indican el nivel de jerarquía. Para este ejemplo se selecciona el término “cernícalos”, y se observa como en la parte inferior del esquema se ocultan los términos asociados a categorías con un nivel de jerarquía mayor. Se resaltan los términos más contiguos al término seleccionado, lo cual permite resaltar aquellos nodos que seleccionamos y ocultar los nodos más extremos, imitando así el efecto de foco y contexto (Pirolli, et al., 2001) descrito en esta sección.

2.5.4 Visualización por categorías

Conocida también como navegación por carpetas, es un tipo de visualización apropiado para el manejo de jerarquías y clasificaciones. Este tipo de visualización se estudió inicialmente con el fin de explorar distintas técnicas de reconocimiento de patrones, por lo tanto pretendía analizar la influencia de los espacios de conceptos y mapas por categorías, con el propósito de resolver problemas de interoperabilidad semántica (Yang, Chen, & Hong, 1999). Su representación gráfica, carece de relaciones visuales a través de enlaces mediante estructuras de tipo padre-

hijo, y permite organizar una gran cantidad de información en espacios reducidos (Bederson, Shneiderman, & Wattenberg, 2002).

Dentro de esta clasificación, se encuentran estrategias de despliegue de información conocidas como facetas (Figura 15 a), donde el mismo grupo de objetos se clasifica en diferentes categorías (Sifer, 2003; Stefaner & Muller, 2007; Tunkelang, 2009). Este tipo de estrategia basado en facetas, surge a partir de los métodos de clasificación implementados en árboles mediante estructuras jerárquicas. Uno de los ejemplos que podemos resaltar en esta categoría se conoce con el nombre de *treemaps* (Shneiderman & Johnson, 1991); el cual es una técnica de visualización que despliega toda la estructura jerárquica sobre una misma área de trabajo, o espacio disponible en pantalla, mediante divisiones rectangulares (Shneiderman, 1992; Shneiderman & Johnson, 1991). Despliega la información de manera jerárquica basándose en la subdivisión recursiva del espacio de trabajo de manera rectangular (van Wijk & van de Wetering, 1999). Un ejemplo lo podemos ver en la Figura 15 b.

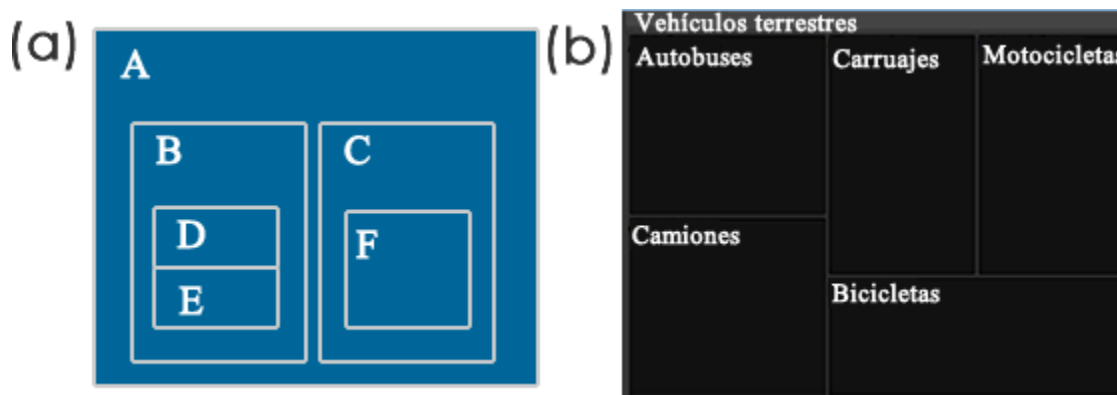


Figura 15. Tipos de representación por categoría (a) Ejemplo de una faceta (b) Adaptación de técnica de visualización treemaps

En la Figura 15 (b), se puede visualizar el tamaño de un rectángulo, el cual es proporcional al tamaño del nodo correspondiente, y los usuarios pueden elegir entre varios algoritmos de particionamiento (Shneiderman, 1992).

Otro ejemplo que se puede identificar dentro de esta categoría se conoce como *icicle* (Kruskal & Landwehr, 1983). Ésta es una técnica que facilita la representación de términos mediante su agrupación jerárquica. Dentro del proceso de navegación se identifican objetos que se unen o abandonan un grupo a medida que se selecciona un nuevo nivel de jerarquía (Kruskal & Landwehr, 1983). En la Figura 16 se puede ver una representación de la interfaz.



Figura 16. Adaptación de técnica de visualización icicle

Dentro de este tipo de visualizaciones, se pueden encontrar una variada lista de propuestas, enmarcadas en clasificar cada una de ellas a partir de su forma, tamaño, propósitos, estructura, etc. La mayoría de ellas se encuentran recopiladas en los trabajos de Draper (2009), Graham (2010) y Gleicher (2011).

Una vez estudiadas la visualización y sus técnicas, así como el concepto de usabilidad, en el siguiente apartado se lleva a cabo una relación de ambos temas.

2.6 Usabilidad de la visualización y navegación de colecciones digitales

Varios autores coinciden (Aula & Käki, 2005; Hargittai, 2004; Hearst, 2009), en que uno de los principios fundamentales para la creación de interfaces de búsqueda es la simplicidad. Por esa razón, es importante analizar la interfaz desde el punto de vista de los usuarios, objetivo principal de IPO, con el fin de maximizar su satisfacción.

Las estrategias de usabilidad dentro del campo de IPO son un factor importante en el desarrollo y la evolución de las interfaces que se encuentran en las bibliotecas y repositorios digitales (Buttenfield, 1999; Fox et al., 1999; Jeng, 2009). Estudios como Marchionini (2006), (White & Roth, 2009) se han centrado específicamente en las estrategias de búsqueda y capacidades de navegación para localizar recursos digitales, demostrando que las estrategias de exploración y búsqueda por sí solas, no son lo suficientemente robustas ni flexibles, para facilitar la localización y acceso a una colección de recursos digitales. En términos de eficiencia, los resultados de estos estudios demuestran que las interfaces visuales a menudo conducen a una subutilización de la información de metadatos muy por debajo de todo su

potencial semántico, a pesar de que estas interfaces visuales son clave para enriquecer el proceso de búsqueda de recursos digitales en un área de conocimiento específico (Cechinel, et al., 2009; J. Fernández, 2001; Pastor, 2009).

En lo relativo a la usabilidad, Hitchcock evaluó la utilidad y la facilidad de la biblioteca digital de acceso libre Citebase^[11], relacionado con los índices de citas en el área de la investigación. Aunque los resultados fueron en general satisfactorios, se detectaron algunos problemas asociados con el nivel de cobertura y de navegación proporcionados por las interfaces. Los mismos resultados se reportaron en un estudio similar sobre la biblioteca eprints E-LIS (Tsakonas & Papatheodorou, 2008), donde la mayoría de los atributos para evaluar la usabilidad fueron criterios altamente valorados por los usuarios, para satisfacer sus necesidades de búsqueda. Por último, cabe destacar los estudios realizados por (Kani, Ghinea, & Chen, 2006), quienes examinaron las preferencias de uso, para la selección de contenidos en tres bibliotecas digitales: la Biblioteca Digital ACM^[12], la biblioteca de referencia técnica de ciencias de computación y redes^[13] y la Biblioteca digital de Nueva Zelanda^[14]. Los autores encontraron que, independientemente del tipo de usuario, la facilidad de aprendizaje y la fiabilidad de las herramientas fueron requisitos fundamentales para la satisfacción de los usuarios en las bibliotecas digitales.

2.6.2. Usabilidad de las taxonomías y esquemas de representación de conocimiento

Los datos de una taxonomía son elementos candidatos para llevar a cabo actividades de exploración mediante técnicas de visualización a partir de su estructura jerárquica. En esta línea, se encuentran varios trabajos que han enfocado sus esfuerzos a analizar las relaciones entre múltiples estructuras taxonómicas. Sin embargo, el diseño de taxonomías que despliegan cientos y miles de nodos, presenta serios problemas asociados con la localización de un gran número de registros (Graham & Kennedy, 2005).

Estudios previos han detallado las ventajas del uso de estructuras taxonómicas para la clasificación de recursos en línea. Morante (2003) presenta una serie de directrices para el desarrollo de taxonomías de sitios Web; resaltando la importancia de la definición de una buena taxonomía, y describiendo buenas prácticas en la aplicación de cada una de ellas. Bajo la misma línea, (Fang & Holsapple, 2007) demostraron a partir de un estudio empírico sobre las estructuras de navegación de sitios web, que las taxonomías pueden beneficiar el acceso a las páginas Web. Los resultados que se obtuvieron demostraron que el diseño de una jerarquía es una estructura de navegación con mayor facilidad de uso, tanto para tareas de exploración simples, como actividades complejas de adquisición de conocimiento. Wang et al., (2008) (Z. Wang, et al., 2008) hallaron sistemas de clasificación y tesauros para representar niveles jerárquicos y clasificar las fuentes de información, y Uddin y Janecek (2007) desarrollaron un método para buscar los documentos a través de taxonomías multidimensionales. De acuerdo con los resultados de las pruebas de usabilidad, encontraron que las taxonomías y clasificación por facetas definidas, eran útiles para la búsqueda y localización de documentos. Finalmente

se resalta el estudio realizado por (Shiri, 2008), quienes encontraron que los elementos de los metadatos y el uso de interfaces visuales bajo las categorías previamente definidas, favorecen los procesos de búsqueda mediante un lenguaje enriquecido de palabras claves para el acceso a una colección de recursos digitales.

Otro estudio representativo realizado por Graham (2000) sobre la exploración de relaciones entre múltiples estructuras jerárquicas en taxonomías botánicas, determinó la utilidad de este tipo de representaciones jerárquicas para la ubicación de términos. Por su parte, (Dinkla, Westenberg, Timmerman, van Hijum, & van Wijk, 2011) llevaron a cabo una comparación de múltiples tamaños de jerarquías en el campo de la biología, y uno de los resultados más representativos del estudio, fue que a través de las taxonomías, los expertos podían descubrir rápidamente correlaciones en un conjunto abundante de microbios a partir del descubrimiento de patrones entre muchas instancias de una jerarquía.

En relación a otros esquemas de representación del conocimiento, existen también un gran número de estudios orientados a la evaluación de interfaces basados en el uso de tesauros. (Blocks, Binding, Cunliffe, & Tudhope, 2002) encontraron que las interfaces de búsqueda basada en tesauros, permitían el uso de funcionalidades básicas sobre usuarios con limitados conocimientos tecnológicos. Sin embargo, la interfaz no facilitaba búsquedas avanzadas basadas en un proceso de exploración para actividades específicas de recuperación de información. Por su parte, (McKay, Shukla, Hunt, & Cunningham, 2004) realizaron la evaluación de una interfaz de búsqueda basada en un tesauro sobre una biblioteca digital, y encontraron que el uso de ventanas múltiples e independientes en los resultados de búsquedas era difícil de manejar por los usuarios. De igual forma, sugieren que las búsquedas se realicen de forma semiautomática, es decir, que el buscador sugiera al usuario términos de búsqueda. (Stafford et al., 2008) evaluaron una versión bilingüe de una interfaz gráfica basada en un tesauro, y encontraron que la integración de funcionalidades de búsqueda y navegación eran útiles para el acceso a recursos digitales. De igual forma identificaron que la interfaz visual fue útil desde el comienzo en que se probó la interfaz, dada que las temáticas que no eran conocidas por los usuarios fueron siendo cubiertas a medida que realizaban el proceso de navegación. Esto les ayudaba a restringir o ampliar los criterios de búsqueda a partir del despliegue de temáticas en su estructura taxonómica.

2.6.3 Revisión de estudios sobre usabilidad de interfaces de búsqueda

Recientemente se han presentado estudios que demuestran la gran aceptación del uso de tesauros en actividades de navegación sobre bibliotecas digitales. Autores como (Shiri, Ruecker, Doll, Bouchard, & Fiorentino, 2011) llevaron a cabo una evaluación de dos interfaces (*Searchlink* y *T-Saurus*) gráficas a partir del uso de tesauros para la búsqueda de recursos digitales multilingües. La interfaz *Searchlink*, proporcionaba una búsqueda lineal y secuencial de recursos digitales mediante facetas; mientras que la interfaz *T-saurus* desplegaba una búsqueda basada en el uso de una estructura de navegación taxonómica por categorías. Uno de los resultados que arrojó el estudio es que la búsqueda por facetas facilitaba más la

localización de recursos digitales que la estrategia usada a partir del despliegue visual por categorías. Sin embargo, los resultados carecen de un estudio preciso de usabilidad que refleje los resultados objetivos del uso de interfaces a partir de actividades de búsquedas específicas.

Por su parte (English, Hearst, Sinha, Swearingen, & Lee, 2002) desarrollaron un framework de interfaz de navegación y búsqueda, caracterizado por la presentación de metadatos jerárquicamente organizados en facetas para guiar a los usuarios hacia posibles opciones de selección y organizar los resultados de búsquedas mediante palabras clave. Como resultado de este desarrollo, se identificó que este tipo de interfaz proporcionaba una manera fácil de ayudar a los usuarios a formular consultas mediante criterios de búsquedas booleanos.

En el ámbito de repositorios digitales, los métodos de búsqueda visuales también han sido estudiados. Algunos de estos métodos se tratan con diferentes enfoques para estudiar las formas de acceso a recursos digitales. El proyecto MACE (Stefaner, et al., 2007) propone varias alternativas de búsqueda visual (semántica, social y contextual) para acceder a los recursos digitales en el "*diseño y la arquitectura*" a través de estrategias de clasificación relacionadas con palabras clave, ubicación, competencia, área social y facetas (Y. Lin, et al., 2010; Stefaner & Muller, 2007). Estos estudios se centraron en las perspectivas de utilizar diferentes rutas de navegación, junto con el etiquetado social para el despliegue de más de 150.000 recursos digitales de 18 agregadores de contenido. El resultado de este estudio sugiere que, en primer lugar, los principios de la navegación con múltiples facetas facilitan los procesos de inmersión a través de actividades de etiquetado de colaboración (Stefaner et al., 2009), y en segundo lugar, que la definición de los metadatos es esencial para mejorar los procesos de búsqueda a través de estrategias de búsqueda contextual.

En ARIADNE (Klerkx, Duval, & Meire, 2004), se evaluaron las ventajas y desventajas (a nivel computacional) de tres técnicas de visualización: diagramas de árbol, hiperbólicas y diagramas de Venn, para acceder a un conjunto de repositorios de objetos de aprendizaje. Esto dio como resultado el desarrollo de un framework que implementaba estas técnicas (Klerkx, Meire, Ternier, Verbert, & Duval, 2005). En este framework, el nivel de acceso fue favorable a partir del uso de interfaces para la búsqueda de objetos de aprendizaje sobre un tema general. Sin embargo, los resultados de las búsquedas específicas a través de las estrategias de clasificación mediante agrupación, no reflejaron los mejores resultados de búsquedas en una colección de recursos digitales.

Cabe destacar que ninguno de estos esfuerzos incluye un estudio a fondo de usabilidad. En su lugar, se centran en las capacidades computacionales de las interfaces en términos de las técnicas aplicadas (Klerkx, et al., 2004), en la identificación de las interfaces con mejores capacidades de navegación (Stefaner, et al., 2007), y en el uso de prototipos de navegación orientados sobre estructuras jerárquicas (Merčun y Žumer, 2009; Merčun, et al., 2012). Por lo tanto, no se ha dado una mayor importancia a la usabilidad y la interacción para la búsqueda eficaz de recursos digitales de acuerdo a criterios como la cobertura temática y búsqueda basadas en estructuras taxonómicas visuales.

Dado que no existen estudios previos de usabilidad acerca de interfaces de usuario adecuadas para la visualización de una colección, exploración y navegación de recursos digitales, en estudios preliminares (Martín, Sánchez, Gaona, & Marianos, 2013), se han encontrado algunos problemas relacionados con el uso de interfaces de búsqueda. El estudio presentó algunos factores que influían sobre los resultados de búsquedas, asociados con la utilidad, facilidad de aprendizaje y conocimiento del área de estudio sobre dos repositorios (Organic.Edunet ^[9] y VOA3R^[10]). Igualmente otros estudios previos de usabilidad sobre la evaluación de técnicas de visualización mediante el uso de estructuras taxonómicas, se ha encontrado algunos problemas relacionados con el proceso de búsqueda de recursos digitales (Gaona, Martín, Feroso, & Sánchez, 2014).

En los siguientes capítulos se llevará a cabo una evaluación de la eficacia y precisión de interfaces de búsqueda visual desarrolladas a partir de ocho técnicas de visualización para el acceso a recursos digitales. El propósito de este estudio es proporcionar a creadores de repositorios digitales la identificación de factores que faciliten el desarrollo e implementación de este tipo de estrategias para la búsqueda de recursos digitales. Dado que el uso de taxonomías presenta desafíos a la hora de gestionar, organizar y visualizar información compleja (Morante, 2003), y dado que el tamaño de los datos es también una cuestión clave para visualizar correctamente la información (Herman, et al., 2000; Huang, Liang, & Nguyen, 2008), hemos tratado de abordar de manera cuidadosa la preparación y el diseño de los experimentos, basados en principios básicos de usabilidad.

2.7 Resumen

En este capítulo se realizó un análisis de la situación actual de las investigaciones y trabajos relacionados sobre el acceso a recursos digitales en repositorios. Por un lado, se presentaron los trabajos más representativos asociados al concepto de esquemas de representación de conocimiento, junto con algunos instrumentos que se utilizan para llevar a cabo este tipo de estrategias de clasificación de recursos digitales. De igual manera se presentaron las diferentes estrategias que se usan en repositorios digitales para llevar a cabo actividades de búsqueda y acceso sobre recursos digitales. A continuación se analizan las características y beneficios del área de visualización, seguido de principios básicos de usabilidad para el diseño de interfaces de búsqueda en repositorios y bibliotecas digitales. Finalmente se llevó a cabo un barrido de las diferentes técnicas desarrolladas en el área de visualización de información, y sus implicaciones para el despliegue de esquemas de representación de conocimiento. El análisis del estado de la cuestión realizado en este apartado, ha permitido identificar las carencias, limitaciones y problemas existentes en repositorios digitales, asociados con el acceso a recursos digitales. Estas limitaciones se abordarán con detenimiento en el siguiente capítulo.

Pie de páginas:

- [1] Disponible online en: <http://209.87.56.111/drupal/>, (último acceso 5 Mayo 2014)
- [2] Disponible online en: <http://www.merlot.org/>, (último acceso 5 Mayo 2014)
- [3] Disponible online en: <http://www.ariadne-eu.org/>, (último acceso 5 Mayo 2014)
- [4] Disponible online en: <http://portal.mace-project.eu/>, (último acceso 5 Mayo 2014)
- [5] Disponible online en: <http://www.organic-lingua.eu/es>, (último acceso 5 Mayo 2014)
- [6] Disponible online en: <http://eprints.ucm.es/>, (último acceso 5 Mayo 2014)
- [7] Disponible online en: <http://www.dspace.org/>, (último acceso 5 Mayo 2014)
- [8] Disponible online en: <http://cds.cern.ch/>, (último acceso 5 Mayo 2014)
- [9] Disponible online en: <http://dspace.mit.edu/>, (último acceso 5 Mayo 2014)
- [10] Disponible online en: <https://ntrs.nasa.gov/search.jsp>, (último acceso 5 Mayo 2014)
- [11] Disponible online en: <http://iplus.ukoln.ac.uk/technology/citebase>, (último acceso 5 Mayo 2014)
- [12] Disponible online en: <http://dl.acm.org/>, (último acceso 5 Mayo 2014)
- [13] Disponible online en: <http://www.ncstrl.org/>, (último acceso 5 Mayo 2014)
- [14] Disponible online en: <http://www.nzdl.org/cgi-bin/library.cgi>, (último acceso 5 Mayo 2014)

Capítulo 3

Planteamiento del problema

El objetivo de este capítulo es ofrecer una síntesis del problema que se pretende resolver. Para ello se abordará una serie de limitaciones que han dado lugar a la identificación de varios problemas relacionados con nuestra área de estudio. En primera instancia se presentan los inconvenientes encontrados en los repositorios digitales para la búsqueda y el acceso a su colección de recursos digitales. A continuación, las limitaciones que los esquemas de representación de conocimiento presentan. Finalmente, se exponen las deficiencias de la definición de metadatos en recursos digitales, como un factor secundario derivado del proceso de investigación realizado, al ser este uno de los factores que afecta también los resultados de búsqueda de recursos digitales de acuerdo a criterios definidos por usuarios.

Capítulo 3: Planteamiento del Problema

"God does not care about our mathematical difficulties. He integrates empirically"
(Albert Einstein)

3.1 Deficiencias de los métodos de búsqueda y acceso sobre objetos de aprendizaje

Para facilitar procesos de gestión y administración, los objetos de aprendizaje son agrupados y almacenados en *repositorios digitales*. Dentro de estos repositorios, podemos identificar dos tipos:

- i) Los que contienen objetos de aprendizaje y sus metadatos. Los objetos de aprendizaje y sus descriptores se encuentran dentro de un mismo sistema.
- ii) Los que contienen sólo los metadatos. El repositorio contiene sólo los descriptores y se accede al objeto a través de una referencia a su ubicación física que se encuentra en otro sistema o repositorio de objetos.

Una búsqueda de objetos de aprendizaje en cualquiera de estos repositorios podría devolver una enorme lista de resultados. Por lo tanto, si no se definen indicadores de calidad que permitan evaluar la calidad de la información registrada, el proceso de búsqueda de objetos de aprendizaje puede resultar una actividad que demande una alta pérdida de tiempo y esfuerzo por parte de un usuario (Kumar, Nesbit, & Han, 2005).

Aun así, el desarrollo de repositorios ha sido una de las iniciativas más generalizadas para centralizar procesos de búsqueda sobre recursos digitales. Se consideran como importantes infraestructuras que facilitan el acceso a colecciones de recursos digitales definidos por museos, bibliotecas, archivos y centros de documentación digital. Estos a su vez, están orientados a realizar actividades relacionadas con el almacenamiento de materiales educativos, con el fin de optimizar procesos de gestión, administración y búsqueda de los mismos. Sin embargo, el gran volumen de recursos digitales ha generado una serie de limitaciones, específicamente aquellos relativos al uso de repositorios para el acceso sobre materiales educativos relevantes. Este tipo de deficiencias ha favorecido el desarrollo y uso de alternativas asociadas con la implementación de *lenguajes enriquecidos y esquemas de representación de conocimiento* para llevar a cabo actividades de clasificación, categorización y gestión de contenidos. La implementación de estas estrategias, ha generado la combinación de una sólida estructura tecnológica, vinculada con una serie de estrategias de enriquecimiento semántico, a partir del uso de esquemas de representación de conocimiento. Estas soluciones, facilitan en cierta medida actividades de gestión y administración por parte de creadores y desarrolladores de repositorios. Sin embargo, para un usuario convencional (estudiantes y/o profesores), el uso de estos repositorios educativos no es sencillo, ya que adolecen de estrategias de acceso a

partir del uso de sus interfaces y mecanismos de búsquedas convencionales (textual, booleano). Este es un factor clave, que puede llegar a dificultar procesos de aprendizaje, generando un progresivo abandono del uso de este tipo de herramientas.

Además de los lenguajes y esquemas de representación del conocimiento, para llevar a cabo un proceso de búsqueda, los repositorios también proporcionan alternativas de acceso, mediante el uso de *interfaces visuales*, si bien algunos repositorios carecen de interfaces que faciliten este tipo de procesos. Bajo esta perspectiva, investigaciones previas han encontrado que algunas interfaces de búsqueda no se ajustan a las necesidades de los usuarios en conjunto (Tenopir, 2003). Con frecuencia, los resultados desplegados no son relevantes de acuerdo a criterios de búsqueda definidos por usuarios (Nash, 2005). Se han identificado problemas de navegación cuando los usuarios desean retornar sobre registros previamente consultados (Jeng, 2005). Por su parte Kim, (2008) encontró una serie de problemas asociados con el diseño de la interfaz de un repositorio institucional en Korea. El estudio demostró que los temas de interés de consulta por parte de los usuarios no eran lo suficientemente visibles, porque los menús de navegación eran demasiado pequeños y el color demasiado oscuro. Otros estudios han permitido revelar limitaciones para combinar estrategias de navegación y métodos de búsqueda (Hartson, Shivakumar, & Pérez, 2004), donde las interfaces no permiten visualizar (a simple vista) el despliegue de materiales disponibles en un repositorio, a partir de un área de conocimiento específico (Hitchcock et al., 2003; Tsakonas & Papatheodorou, 2008). Esta última limitación resulta ser un factor determinante para la continuidad de un usuario a través del uso de estas herramientas, dado que les resulta difícil determinar si vale la pena continuar con el proceso de exploración de materiales en el repositorio, o si es mejor valerse de otras estrategias de búsqueda externas.

A los análisis mencionados anteriormente sobre las interfaces de acceso a los repositorios, que como se ha comprobado ponen de manifiesto algunas de sus limitaciones, se han añadido *estudios de usabilidad* sobre estos sistemas de acceso. Estos estudios han identificado una serie de limitaciones que presentan algunas bibliotecas y repositorios digitales. (Tsakonas & Papatheodorou, 2007) realizaron un análisis de usabilidad sobre la biblioteca digital E-LIS, encontrando dificultades para la realización de ciertos procesos de búsqueda y generando un alto consumo de tiempo para llevar a cabo ciertas tareas de búsquedas y grandes esfuerzos para comprender el uso las interfaces. Por su lado (Marchionini, 2006; White & Roth, 2009), resaltan el uso de las estrategias de búsquedas exploratorias y las capacidades de localización de recursos digitales. Los autores, resaltan la importancia de complementar este tipo de estrategias, dado que por sí solas no son lo suficientemente robustas y flexibles para facilitar la localización y acceso a una colección de recursos digitales. Bajo la misma línea, se encuentran los estudios realizados por (Buchanan & Salako, 2009; Petrelli, 2008), donde identificaron un alto consumo de tiempo por parte de usuarios para realizar procesos de búsqueda a partir de una limitada lista de criterios de filtrado por parte de la herramienta. De la misma forma, mencionan dentro de los resultados obtenidos algunas dificultades por parte de los usuarios, quienes finalmente indicaron que las interfaces eran muy complejas para realizar actividades de búsqueda, dado que no lograban ubicarse con facilidad dentro de las estructuras de navegación.

Otros estudios igualmente relacionados con el análisis de la usabilidad en el acceso a recursos de bibliotecas digitales, son por ejemplo el llevado a cabo por Klerkx (2004). En este trabajo, mediante un estudio de usabilidad, encontraron que los criterios de búsquedas a partir de técnicas de agrupación no reflejaban los mejores resultados de búsquedas dentro de una colección de recursos digitales. Por su lado Xie 2006; Tsakonas 2007 y Deng 2009 mencionan una serie de frustraciones por parte de usuarios asociadas con el uso de interfaces de búsqueda; factores que junto con procesos previos de autenticación, debían realizar los usuarios para llevar a cabo este tipo de actividades. Otros estudios recientes de usabilidad, encontraron serias dificultades para realizar actividades de búsqueda y acceso sobre recursos digitales relevantes a partir del uso de sus interfaces. (Khoo, Kusunoki, & MacDonald, 2012), realizaron un estudio de usabilidad sobre la biblioteca digital Internet Public Library^[1] mediante la participación de usuarios expertos en usabilidad y técnicas derivadas de “*thinking aloud*” y “*eye-tracking*”. Como resultado de este estudio, lograron identificar algunas limitaciones asociados con la localización de recursos digitales a partir del uso de sus interfaces. De igual forma, los usuarios manifestaron frustraciones debido al elevado número de resultados imprecisos obtenidos a partir de los criterios de búsqueda definidos.

Sobre el área de repositorios académicos, también se resaltan algunos resultados de estudios relacionados a partir de criterios de usabilidad. Kim (2008), encontró una serie de limitaciones asociadas con el uso de interfaces de búsqueda y las estrategias de despliegue de recursos digitales. Los autores proponen, una lista de sugerencias y recomendaciones relacionadas con la mejora en la definición de criterios de búsqueda y la distribución de resultados en pantalla. Finalmente, cabe mencionar algunos de los resultados obtenidos de estudios recientes asociados con el uso de interfaces en repositorios académicos. En Martín et al. (2013) se identificaron deficiencias en las interfaces por parte de usuarios, relacionadas con la utilidad, aprendizaje y conocimiento del esquema de clasificación. En Gaona et al. (2014), se identificó que la mala definición de un esquema de representación de conocimiento es uno de los factores que dificulta el proceso de búsqueda de recursos digitales mediante el uso de interfaces a partir técnicas de visualización; factores que impiden la navegación sobre estructuras de representación de conocimiento para localizar recursos digitales.

En términos de eficiencia, los resultados de estos estudios demuestran que las interfaces de búsqueda en ciertas condiciones conducen a una subutilización de información, por ejemplo resultados de procesos de búsqueda no relevantes de recursos digitales asociados a otras temáticas o áreas de conocimiento. Resultados que suelen asociarse a partir de una mala definición de sus metadatos, perdiendo toda capacidad semántica para realizar búsquedas enriquecidas, basadas en temáticas o algún área de conocimiento específico (Cechinel, et al., 2009; J. Fernández, 2001; Pastor, 2009).

3.2 Limitaciones de los esquemas de representación de conocimiento en repositorios

Con el rápido crecimiento que han tenido las colecciones de recursos digitales en los últimos años, las organizaciones han desarrollado diversas estrategias y políticas para fomentar y facilitar su uso (Liew & Foo, 1999; Prabha, 2007; S. Zhang, Shen, & Ghenniwa, 2004). Este crecimiento, en el caso en particular del sector de la educación, se ha convertido en parte esencial dentro de un entorno de aprendizaje (Armstrong et al., 2001; Hayati & Jowkar, 2008; Nicholas, Huntington, & Jamali, 2007; Pancheshnikov, 2007). Sin embargo, hay opiniones generalizadas que indican que este tipo de recursos orientados hacia la educación no se utilizan en todo su potencial con el fin de mejorar procesos de enseñanza e investigación (Fox & Logan, 2005; Jain & Babbar, 2006; Shuling, 2007). A esta incapacidad se le atribuyen una serie de factores asociados con falta de: i) competencia, ii) formación, y iii) falta de tiempo para el uso de los mismos (Roberts, 1995; Shuler, 2007; ur Rehman & Ramzy, 2004). Específicamente sobre este último factor, se han realizado grandes esfuerzos y propuestas para el desarrollo de estrategias que permitan clasificar y organizar recursos digitales, los cuáles se han visto materializados a partir del uso de esquemas de representación de conocimiento (tesauros y ontologías). La aplicación de este tipo de estrategias ha sido de gran provecho para personas expertas en actividades relacionadas con la gestión y administración de un repositorio digital (creadores y administradores). Sin embargo, estos esfuerzos no se han enfocado sobre las necesidades específicas de acceso por parte de usuarios para localizar recursos digitales (estudiantes, profesores). Bajo este enfoque, a continuación se presenta una serie de trabajos relacionados donde se han encontrado inconvenientes asociados con el uso de instrumentos de representación de conocimiento por parte de usuarios finales.

Sutcliffe et al. (2000), llevaron a cabo una evaluación de la interacción de los usuarios a partir del uso de un tesoro, y encontraron que los mejores resultados obtenidos fueron por parte de usuarios investigadores, quienes utilizaban la interfaz de manera eficaz y dedicaban un mayor tiempo de uso para actividades relacionadas con sus actividades de trabajo. Sin embargo, en algunas oportunidades presentaron resultados deficientes, atribuidos a la falta de motivación, y también, por el mal uso del sistema de visualización propuesto. (Shiri & Revie, 2005) plantearon un estudio de usabilidad sobre interfaces de búsqueda a partir del uso de tesauros. Encontraron que la mayoría de los usuarios consideraban que la navegación del tesoro a través de un vocabulario controlado era fácil de usar. Sin embargo, el proceso de búsqueda y navegación del tesoro mediante términos preferidos y no preferidos generaron algunas confusiones sobre el uso de palabras en algunas tareas de búsqueda; inconvenientes que se reflejaron a partir de otros problemas de usabilidad identificados en las interfaces. Este tipo de problemas a nivel terminológico, es otro de los aspectos que a menudo se atribuyen a las dificultades que enfrentan tanto creadores, como usuarios de un repositorio digital definido a partir de un tesoro (Aitta, Kaleva, & Kortelainen, 2008). Esta característica asociada a un esquema de representación de conocimiento, en algunas ocasiones limita el desarrollo de mecanismos de búsqueda; dado que si no hay un dominio completo del esquema de representación de conocimiento propuesto, la vinculación de una colección de recursos

digitales sería una actividad difícil de llevar a cabo por parte de cualquier mecanismo de búsqueda.

Respecto a los estudios asociados con el uso de ontologías, algunas investigaciones indican que este tipo de instrumentos de representación de conocimiento presentan una serie de inconvenientes para el acceso sobre recursos digitales (Gašević & Hatala, 2006; Li, et al., 2005). Dentro de los problemas más frecuentes, se encuentra la preferencia de los usuarios por el uso de terminologías comunes. Esto normalmente se ve reflejado en la falta de comprensión de los usuarios para entender la estructura ontológica. El uso de clases en la definición de las ontologías para representar conceptos tiene varias desventajas. Una de ellas, se encuentra asociada con la rigidez del esquema y las complicaciones a nivel de razonamiento (Gašević & Hatala, 2006). Otros estudios han identificado errores asociados a la definición de la taxonomía de una ontología (Gómez, 1999; Noshairwan, Qadir, & Fahad, 2007). Gómez et al. (1999) identificaron y categorizaron tres tipos de errores que se encuentran relacionados con la definición de una ontología, asociados con i) la inconsistencia, ii) la completitud, y iii) la redundancia de información. La inconsistencia hace referencia a que la ontología pueda tener algún tipo de información contradictoria. La completitud se refiere a que los conceptos asociados a la ontología, no se encuentran completamente definidos. La redundancia hace mención a que la misma información es referida por la ontología más de una vez. Estos resultados reflejan, que el uso de este tipo de instrumentos debe tener una serie de criterios previamente establecidos tanto para su diseño e implementación a nivel semántico, como para su integración a partir del diseño de interfaces efectivas, que faciliten procesos de búsqueda de recursos digitales por parte de estudiantes y profesores.

Todo lo anterior pone de manifiesto las serias limitaciones de los esquemas de representación del conocimiento, tesauros y ontologías, a la hora de facilitar el acceso a los recursos digitales, especialmente para los usuarios finales que desean reutilizar dichos recursos.

3.3 Limitaciones en el uso de metadatos

Analizadas las limitaciones de los principales esquemas de representación del conocimiento, otro aspecto relacionado para mejorar los procesos de búsqueda y acceso a recursos digitales son las descripciones de metadatos asociadas a dichos recursos.

Los repositorios digitales son importantes infraestructuras que facilitan el acceso a colecciones de recursos digitales de museos, bibliotecas y archivos. Desafortunadamente, como se acaba de mencionar en los apartados anteriores, se han identificado algunas deficiencias asociadas con la búsqueda y localización de recursos digitales, de acuerdo a criterios de búsqueda por parte de un usuario; se trata de factores que limitan el uso generalizado de los valiosos servicios que ofrecen. A pesar de que la mayoría de los repositorios incluyen sistemas de clasificación de conocimientos (por ejemplo, las ontologías, tesauros) para ayudar a gestionar el contenido por parte de expertos y creadores de un repositorio digital, la falta de mecanismos

efectivos restringen a usuarios para localizar recursos digitales de acuerdo a sus criterios de búsqueda.

Por ello otra de las estrategias que se han definido para mejorar los procesos de búsqueda de un recurso digital, se encuentra asociada con la descripción de metadatos. La creación de un metadato es una actividad que se lleva a cabo de manera manual (una persona) o de forma automática (un ordenador). Esto ha generado una serie de inconsistencias y limitaciones a la hora de gestionarlos, debido a los complejos procesos de actualización que se definen a través de estructuras jerárquicas y semánticas (esquema de representación de conocimiento), a partir de instrumentos como:

- El uso de vocabularios controlados (ej. tesauros)
- La aplicación de un dominio de uso específico (ej. ontologías)
- La aplicación de un dominio de conocimiento de uso libre (DBpedia).

El éxito de la localización de recursos digitales, depende en gran medida de la calidad con que los metadatos se han diseñado. Este factor es una condición esencial para obtener resultados de búsquedas relevantes a partir del uso de motores de búsqueda definidos en un repositorio (De la Prieta & Gil, 2010; Muñoz, Calvillo, Ochoa, Santaolaya, & Álvarez, 2010); por otro lado, la calidad de los metadatos es importante para mejorar las estrategias de indexación de objetos de aprendizaje almacenados en ellos (Ochoa, et al., 2005; Stuckenschmidt, et al., 2004; Wiley, 2002a). Por lo tanto, la omisión de metadatos afecta negativamente los resultados de búsqueda a partir de criterios de búsqueda definidos por ejemplo dentro de un área de conocimiento específico.

La calidad de los contenidos es un indicador que permite evaluar los recursos digitales. Sobre este escenario de trabajo, existen varios estudios que se refieren al diseño de métodos para evaluar la calidad de los recursos digitales (Chuanjun, 2004; Downes, 2007; Gonçalves, Moreira, Fox, & Watson, 2007; Sanz, et al., 2009), así como a la calidad de los contenidos existentes en las colecciones de recursos digitales (Chuanjun, 2004; Downes, 2007). A partir de estos métodos, se definen una serie de criterios de valoración para llevar a cabo evaluaciones de calidad de metadatos sobre una colección de recursos digitales. Sin embargo, algunos autores como como (Sicilia, Sánchez, & Benito, 2006), señalan que no existe un patrón que permita definir criterios de calidad de metadatos para implementarlo sobre cualquier escenario de trabajo.

El problema de la baja calidad de los metadatos en repositorios digitales ya se ha mencionado en otras investigaciones. (Shreeves et al., 2005) estudiaron el caso de una gran biblioteca académica que contenía 14.425 registros, junto con otra pequeña biblioteca académica que contenía 1.599 recursos. Ambas utilizaban un marco de calidad de la información disponible para evaluar la calidad de la información y se enfocaron principalmente en la integridad de los registros de metadatos y la coherencia estructural / semántica. Tomando un conjunto básico de

ocho elementos de Dublin Core (DC), los autores midieron la integridad y encontraron que la primera colección contenía un 71% de los registros incompletos, mientras que el segundo contenía 43% de los registros incompletos, lo que significaba que no disponían por lo menos uno de los ocho elementos obligatorios definidos por Dublin Core. De la misma manera Barton et al. (2003) exponen algunas consideraciones de áreas donde surgen problemas relacionados con la definición de metadatos, algunos de ellos asociados con: 1) la ortografía y las abreviaturas; 2) otros campos asociados con el proveedor del contenido; 3) campo de autor, 4) campos asociados con la temática, y 5) la descripción de la fecha.

3.4 Resumen

En este capítulo se han descrito las limitaciones e inconvenientes más representativas sobre el acceso de recursos digitales, dentro de los cuáles se pueden ver varios desafíos que se concretarían en varios escenarios de trabajo, a saber:

1. Diseño instruccional, mediante el uso de herramientas que favorezcan el desarrollo de material didáctico basado en especificaciones y estándares
2. Reutilización de recursos, a partir de material educativo abierto que permita su libre distribución en diferentes entornos tecnológicos y ambientes de aprendizaje
3. Diseño de interfaces de búsquedas apropiadas, que facilitarían en gran medida la accesibilidad sobre una colección de recursos digitales para identificar material educativo relevante, asociados a partir de criterios de búsqueda definidos por un usuario.
4. Esquema de representación de conocimiento acorde a la colección de recursos digitales alojados en un repositorio.
5. Enriquecimiento semántico, mediante el uso de metadatos de calidad que faciliten su descripción y una oportuna localización a partir de criterios de búsquedas definidos por las descripciones más representativas.

Para propósitos de esta investigación, se abordarán los tres últimos factores, los cuales atañen directamente al tema de estudio. Estos, factores se desglosarán en el siguiente capítulo mediante el diseño de una solución orientada a facilitar a los creadores de repositorios digitales, ofrecer alternativas de búsquedas visuales que faciliten el acceso de usuarios sobre una colección de recursos digitales a partir del uso de técnicas de visualización.

Pie de página:

[1] Disponible online en: <http://www.ipl.org/> (último acceso 5 Mayo 2014)

Capítulo 4

Diseño de una solución

Este capítulo presenta las diferentes fases de trabajo que dieron lugar al análisis y diseño de interfaces de búsqueda gráficas para el acceso a recursos digitales, a partir de técnicas de visualización. El propósito de este capítulo es estudiar todas las alternativas de las estrategias de visualización existentes, e identificar aquellas que faciliten la búsqueda y el acceso a recursos digitales a partir de criterios de evaluación que permitan someterlas a prueba (Capítulo 6). Por lo tanto, en la primera sección de este capítulo se describen las diferentes fases metodológicas de trabajo. Para llevar a cabo este estudio, se tomó como caso de estudio la biblioteca digital Europea^[1] con el propósito de explorar la cobertura de recursos digitales sobre un área de conocimiento específico. A continuación se realiza un análisis de calidad de metadatos preliminar de los recursos digitales explorados, con el fin de determinar criterios de búsquedas candidatas para aplicarlas a través de técnicas de visualización. En la segunda sección, se analiza la estructura taxonómica a partir de las áreas de conocimiento seleccionadas, para llevar a cabo el diseño de un esquema de navegación. De igual forma se presenta el proceso de transformación de datos y la vinculación de recursos digitales asociados a cada una de los nodos representados a través del esquema de representación de conocimiento diseñado. Finalmente, se presenta la integración del esquema de representación de conocimiento planteado, y una propuesta de arquitectura para la búsqueda de recursos digitales a partir de interfaces gráficas.

Capítulo 4: Diseño de una solución

"If I can't picture it, I can't understand it"
(A. Einstein)

4.1 Modelo de trabajo

Para realizar el análisis y diseño de interfaces de búsqueda a partir de técnicas de visualización, se planteó inicialmente un caso de estudio para la exploración de recursos digitales sobre la biblioteca digital Europea asociados a una rama de conocimiento del tesoro de Arte y Arquitectura AAT(Gaona, et al., n.p.). Este estudio permitió identificar inicialmente, en qué medida Europea cubría ciertos temas o áreas de conocimiento para llevar a cabo procesos de reutilización y desarrollo de objetos de aprendizaje en áreas de conocimiento asociados al patrimonio cultural europeo. Por otro lado, se analizó la relación de algunos elementos de metadatos de estos recursos digitales, para verificar la integridad de los resultados de búsqueda de los usuarios, y la disponibilidad de los recursos digitales para acceder a ellos. Uno de los resultados de este proceso, logró vislumbrar una carencia en la definición de metadatos de la mayoría de recursos digitales explorados en Europea. Resultado que permitió identificar una serie de deficiencias en los procesos de búsquedas de recursos digitales a partir de un área de conocimiento. A partir de estos resultados, se definieron una serie de actividades asociadas al diseño de alternativas de búsqueda gráficas a partir de un esquema de representación de conocimiento, mediante el uso de técnicas de visualización. La Figura 17 resume la ruta de trabajo que dio origen al desarrollo de esta propuesta.

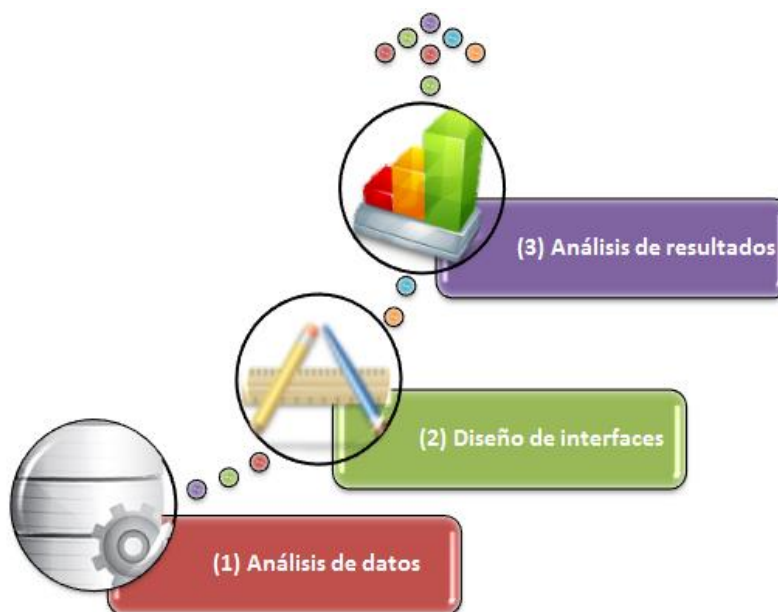


Figura 17. Modelo de trabajo planteado

Prácticamente el modelo de trabajo se puede resumir en las siguientes fases de trabajo, donde de manera preliminar podemos identificar en la fase de *análisis de datos*, una serie de actividades relacionadas con: 1) los criterios de selección del área de conocimiento a explorar en Europeana para determinar la cobertura temática, 2) proceso de exploración y extracción de metadatos, 3) análisis de cobertura de recursos digitales, y 4) criterios de selección de metadatos para la definición de estrategias de búsqueda visual. La fase de *diseño de interfaces* comprende tres actividades: 1) análisis y definición de esquema de representación de conocimiento, 2) análisis y selección de técnicas de visualización y 3) análisis y definición de bibliotecas de funciones gráficas y herramientas de visualización. Finalmente en la fase de *análisis de resultados*, se presenta el diseño de un caso de estudio de usabilidad, y los resultados del análisis de calidad de metadatos que dieron lugar a la selección de metadatos para definir criterios de búsqueda visual. Estos resultados se describen en mayor detalle en la Sección 5.3.

A continuación se describen en profundidad la definición de estas fases junto con cada una de las actividades planteadas.

4.2 Análisis de datos: Caso de estudio biblioteca digital Europeana

Para llevar a cabo un estudio real sobre el uso interfaces de búsquedas y su relación con una colección de recursos digitales, se tomó como caso de estudio la biblioteca digital Europeana. Se ha tomado Europeana como caso de estudio, dado que presenta un conjunto de características relevantes: i) un modelo semántico de intercambio de datos EDM (Doerr et al., 2010), ii) gran número de recursos digitales asociados al área de patrimonio cultural europeo (más de 20 millones de recursos digitales a la fecha), iii) el mayor grupo de proveedores de contenidos a nivel europeo, y finalmente iv) posibilidad de reutilización de recursos digitales abiertos. Por tanto es una iniciativa con una gran oportunidad para llevar a cabo actividades de reutilización de recursos digitales, a fin de permitir a profesores o profesionales dentro del patrimonio cultural, crear objetos de aprendizaje para fines educativos.

A partir de estas características, se llevó a cabo un proceso de exploración de recursos digitales de Europeana para determinar su nivel de cobertura a partir de una rama de conocimiento definido por un tesoro de Arte y Arquitectura AAT. Para realizar este proceso se identificaron cuatro actividades dentro de las cuales destacamos: 1) Proceso de análisis y selección del área de conocimiento a explorar, 2) proceso de exploración y extracción de recursos digitales, 3) análisis de cobertura de recursos digitales explorados, 4) definición de criterios de búsqueda a partir de un análisis de sus metadatos.

4.2.1 Análisis y selección del área de conocimiento a explorar

Esta actividad tenía como propósito identificar el área de conocimiento para llevar a cabo el análisis de cobertura de recursos digitales sobre Europeana. El término de cobertura (o cobertura temática) se ha adaptado de definiciones previas realizadas por Codina (2000) y (Whitehall, 1992, 1995), para referirnos al número de recursos digitales asociados a una serie de temas o áreas de conocimiento. Así, se analiza la cobertura con el fin de explorar la profundidad de la biblioteca digital Europeana en términos de áreas de conocimiento con un mayor número de recursos digitales. Por lo tanto, para el estudio de cobertura, se tomó como base un conjunto de términos asociados a las temáticas de “*estilos y periodos*”, específicamente las ramas de conocimiento “*estilos y períodos por era*” y “*estilos y períodos por región*” del tesauro de Arte y Arquitectura. A continuación se presentan una serie de criterios que se tomaron en cuenta para la selección del tesauro AAT:

- Su representación amplia y detallada de estructura jerárquica (Tudhope, Binding, Blocks, & Cunliffe, 2006).
- El dominio de conocimiento (arte y arquitectura) del tesauro AAT es uno de los más completos y ampliamente reconocidos (Aitchison, Gilchrist, & Bawden, 2001).
- Tiene la posibilidad de múltiples descriptores de conceptos, permite la indexación y consultas muy amplias y precisas (Soergel, 1995).

Por otro lado, el dominio de conocimiento fue seleccionado de acuerdo a una serie de factores asociados por:

- Los temas definidos en la rama de conocimiento de “*estilos y periodos*” del tesauro AAT, tienen una relación directa con la temática de los recursos digitales de Europeana respecto al patrimonio cultural.
- Los términos de AAT están relacionados con un conjunto de vocabularios estructurados relacionados con el arte y la cultura material (Weinberg, 1995).
- El tesauro AAT presenta un esquema de representación de conocimiento maduro (Tudhope, et al., 2006).

Una vez identificado el dominio de conocimiento a explorar sobre la biblioteca digital Europeana, fue indispensable conocer el modelo de representación de datos de Europeana para identificar los elementos de metadatos definidos en cada uno de los recursos digitales para realizar el proceso de extracción. Por tal motivo se tomó como referencia el modelo de intercambio de datos definido por Europeana, conocido en inglés con el nombre de Europeana Data Model EDM (Doerr, et al., 2010). Este modelo de intercambio de datos tiene como objetivo principal preservar los metadatos originales definidos por los proveedores de contenidos, y ser un modelo con una amplia capacidad de expresión semántica. Su definición se basa en las

buenas prácticas de la Web semántica, compatible con los estándares como el protocolo Open Archives Initiative Object Reuse and Exchange (OAI-ORE) (Lagoze et al., 2007) para el intercambio de recursos de Internet y la interoperabilidad. El modelo también utiliza lenguajes de descripción semántica a través de lenguajes como Resource Description Framework (RDF)^[1] y Simple Knowledge Organization System (SKOS)^[2], este último de gran aceptación para la representación de estructuras básicas y contenidos de esquemas conceptuales mediante el uso de taxonomías, tesauros y cualquier tipo de vocabulario controlado.

EDM mantiene la definición de un conjunto de metadatos, definidos por su anterior modelo de representación de datos, Europeana Semantic Elements ESE (Clyphan & Sugimoto, 2009). Por lo tanto, mantiene un conjunto de 28 elementos de metadatos, de los cuales 15 están definidos por el estándar Dublin Core (DC) (Dublin Core, 2008), un subconjunto de los 22 términos de DC. El resto de metadatos, son definiciones creadas por Europeana a fin de cumplir con necesidades propias del proyecto, asociados con criterios de búsqueda. La Figura 18 describe estos elementos, definidos previamente en el anterior modelo de intercambio de datos ESE (Clyphan, 2009), clasificados por elementos obligatorios, recomendados y opcionales.



Figura 18. Metadatos presentes en el modelo de datos de Europeana EDM

La Figura 18 presenta tres categorías de metadatos definidos en EDM. El primer conjunto de metadatos son obligatorios, por lo tanto son descripciones que tanto los proveedores de contenidos, como Europeana, asignan de manera obligatoria. El segundo conjunto de metadatos son recomendados, y permiten complementar algunas descripciones de metadatos definidos en el grupo de elementos obligatorios. Finalmente el tercer conjunto de metadatos opcionales, se encuentran asociados a descripciones adicionales del recurso digital, como la ubicación, la versión, características del formato, etc. Dada la gran cantidad de elementos de metadatos, el estudio de calidad de metadatos definido en la Sección 4.2.3, se llevó únicamente a cabo para los dos primeros grupos: obligatorios y recomendados.

4.2.2 Proceso de exploración y extracción de recursos digitales

Posterior a la selección de la rama de conocimiento y la selección de metadatos de los recursos digitales definidos a través del modelo de representación de datos de Europeana EDM, fue necesario determinar una estrategia para vincular los recursos digitales de Europeana asociados a un área de conocimiento a partir de los términos seleccionados por el tesaurus AAT. Por este motivo, se desarrolló un Web Crawler (Gaona, Sánchez, & Gaona, 2013) con el fin de llevar a cabo un proceso de exploración de los recursos digitales descubiertos en Europeana por cada término del tesaurus AAT.

La estrategia de los Web Crawler consiste en analizar la sintaxis de una página Web y extraer la información de acuerdo a su estructura (Tadapak, Suebchua, & Rungsawang, 2010). Este tipo de estrategia de extracción fue seleccionado a partir de una serie de características definidas en estudios preliminares sobre este tipo de herramientas. Estas características se encuentran asociadas con una sólida definición de su arquitectura (Baeza & Castillo, 2004; Tripathy & Patra, 2008), son herramientas altamente escalables (Boldi, Codenotti, Santini, & Vigna, 2004), favorecen actividades de optimización de procesos (Edwards, McCurley, & Tomlin, 2001) y presentan altos niveles de eficiencia en procesos de extracción de contenidos (C. Castillo, 2005).

El Web Crawler fue desarrollado en Java mediante el uso de Netbeans como entorno de desarrollo. Se utilizaron estrategias de análisis y parseo de datos definidos a través de la librería Jerico^[3]. Esta biblioteca de funciones contiene métodos que permiten explorar un sitio Web para identificar etiquetas de consulta, obtener resultados, guardarlos en memoria, etc. La Figura 19, presenta el modelo de extracción de datos desarrollado.

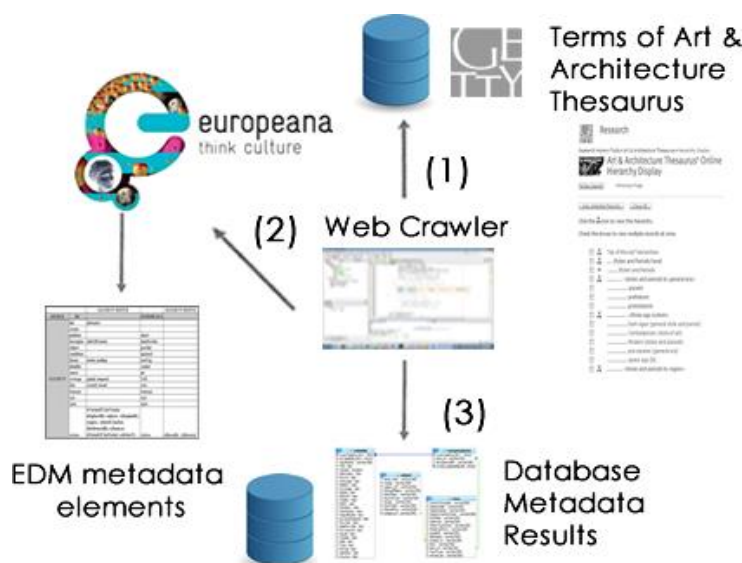


Figura 19. Modelo de extracción de datos definido para el Web Crawler

El modelo de extracción presentado en la Figura 19 se resume a continuación:

- 1) El Web Crawler realiza una consulta para explorar y seleccionar el conjunto de términos asociados a las áreas de conocimiento “*estilos y períodos por era*” y “*estilos y períodos por región*” del tesoro AAT. Para nuestro caso de estudio se realizó una selección de 118 términos, los cuales se encuentran almacenados en una base de datos. Los criterios de selección de los términos se detallan en la Sección 4.3.1.
- 2) Una vez el Web Crawler identifica el número de términos a explorar, toma el primer término alojado en la base de datos del tesoro AAT y lanza un proceso de consulta sobre la biblioteca digital Europeana, para identificar el número de recursos digitales asociados a este término en Europeana.
- 3) Finalmente, mediante un proceso de exploración a través del mecanismo de paginación de Europeana, recorre todos los recursos digitales encontrados y procede a extraer los metadatos definidos por cada recurso digital, para su posterior almacenamiento en una base de datos. En la Figura 20 se presenta un ejemplo de un recurso digital explorado y los metadatos asociados al mismo.

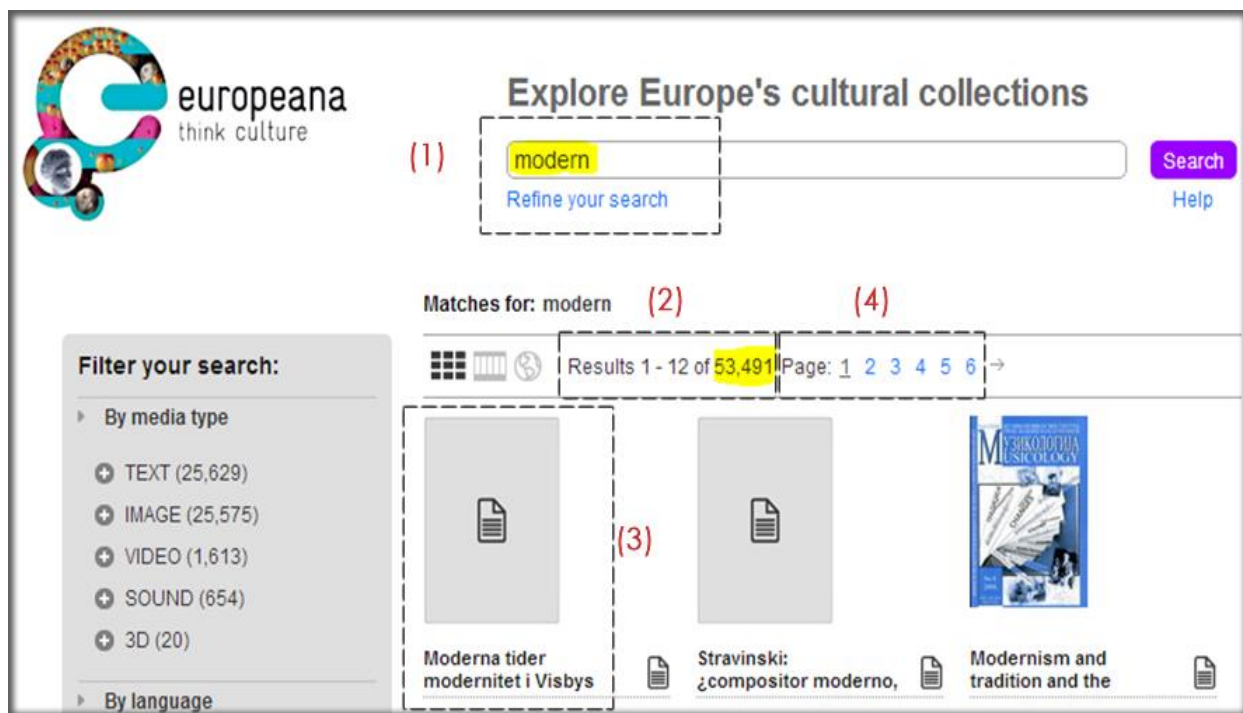


Figura 20. Modelo para la extracción de datos definidos en el Web Crawler

La Figura 20 presenta un ejemplo del proceso de extracción que realiza el Web Crawler: 1) el Web Crawler toma el término “*modern*” del tesoro AAT y lanza una consulta. 2) dentro del proceso de búsqueda se obtienen una serie de recursos en Europeana (53.491 en este caso), este dato es almacenado en la base de datos. 3) la herramienta toma el primer recurso digital, realiza un recorrido de los metadatos definidos en el mismo para realizar un proceso de almacenamiento en la base de datos. (4) posteriormente el Web Crawler continúa con el

recorrido de todos los resultados arrojados a partir del método de paginación definido por Europeana.

Los metadatos extraídos de cada recurso digital son conformes con los elementos definidos en el modelo de datos de Europeana EDM (Figura 18). En la Figura 21, se muestra un ejemplo de la extracción de metadatos de uno de los recursos digitales. Este recurso presenta 10 elementos de metadatos que describen algunas de las características más representativas del recurso digital:

- *Title*: Nombre que se da al recurso.
- *Creator*: Entidad encargada principalmente de crear el recurso.
- *Contributor*: Una entidad responsable de hacer contribuciones al recurso.
- *Date*: fecha de creación del recurso.
- *Type*: El tipo del recurso original o el objeto digital según lo registrado por el titular de contenido, este elemento suele incluir valores como fotografía, pintura, escultura, etc.
- *Description*: Una descripción del recurso original o de forma digital.
- *Data provider*: El nombre o identificador de la organización que aporta datos a Europeana.
- *Provider*: Nombre de la organización que entrega datos a Europeana.
- *Identifier*: Referencia inequívoca al recurso en un contexto dado.
- *Format*: El formato, medio o las dimensiones del archivo de recurso físico.
- *Language*: El idioma del recurso.

The screenshot displays a digital resource page from Europeana. On the left, there is a placeholder for a document icon. The main content area on the right is titled "Moderna tider modernitet i Visbys bouppteckningar och press 1880-1910". Below the title, various metadata fields are listed: **Creator**: Ahlman Robert 1972- ; Höskolan på Gotland, Avdelningen för Historia | ▶; **Contributor**: Sandström Åke ; Höskolan på Gotland Avdelningen för Historia; **Date**: 2001 ; 2009-08-03T13:43:00; **Type**: studentThesis | ▶ Student thesis | ▶ Studentuppsats (Examensarbete) | ▶ Oppgave | ▶ book | ▶ text | ▶ Subject/course | ▶; **Description**: 10 poäng / 15 hp. A paragraph of text follows: "It is a difficult task to define the little word "modernity". One may argue quite a lot about what modernity is as well as when modernity breaks trough. The burgess class was the group who initiated modernity in most cases. Since these people". Below this, there is a "See more" link. Further down, **Data provider**: Uppsala Universitet | ▶ and **Provider**: The European Library | ▶ Sweden | ▶ are listed. At the bottom, there is an "Explore further!" section with a "Similar content" button and a row of five small image thumbnails.

Moderna tider modernitet i Visbys bouppteckningar och press 1880-1910

Creator: Ahlman Robert 1972- ; Höskolan på Gotland, Avdelningen för Historia | ▶

Contributor: Sandström Åke ; Höskolan på Gotland Avdelningen för Historia

Date: 2001 ; 2009-08-03T13:43:00

Type: studentThesis | ▶ Student thesis | ▶ Studentuppsats (Examensarbete) | ▶ Oppgave | ▶ book | ▶ text | ▶ Subject/course | ▶

Description: 10 poäng / 15 hp

It is a difficult task to define the little word "modernity". One may argue quite a lot about what modernity is as well as when modernity breaks trough. The burgess class was the group who initiated modernity in most cases. Since these people

See more ▶

Data provider: Uppsala Universitet | ▶

Provider: The European Library | ▶ Sweden | ▶

Explore further!

Similar content

Figura 21. Metadatos extraídos de un recurso digital en Europeana

4.2.3 Análisis de cobertura de recursos digitales explorados

A través del proceso de exploración el Web Crawler identificó inicialmente cerca de (556.000) recursos digitales asociados a los 118 términos seleccionados del tesauro AAT. Sin embargo, con el apoyo de un grupo de expertos en patrimonio cultural, se llevó a cabo una revisión de los resultados explorados y se encontró que la mayoría de los recursos explorados no correspondía al dominio de conocimiento definido por los términos del tesauro AAT. Por esta razón, se diseñó una estrategia de exploración mediante búsquedas refinadas y palabras claves relacionados al dominio de conocimiento de cada término del tesauro AAT. Esta estrategia condujo a resultados de mayor relevancia de acuerdo al área de conocimiento de cada uno de los términos del tesauro. El número de recursos digitales fue aproximadamente (44.280) en relación a estas áreas de conocimiento. Estos resultados se presentan en mayor detalle en el Anexo A, donde se ilustran los resultados del primer proceso de exploración (normal), vs. resultados mediante el uso de palabras claves asociados a los términos AAT(refinada). La Figura 22 presenta los resultados obtenidos de la cobertura de Europeana, asociados con el área de conocimiento de “*Styles and Periods*” del tesauro AAT, mediante esta última estrategia de extracción.

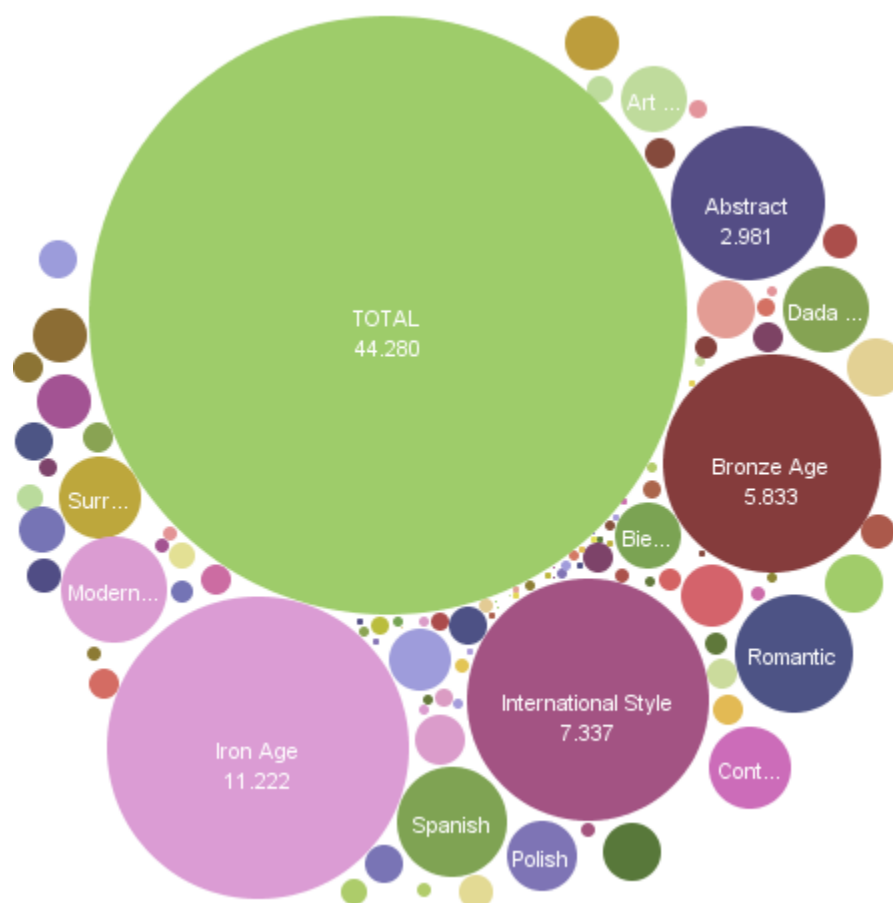


Figura 22. Número de recursos digitales asociados a ciertas temáticas del tesauro

La Figura 22 presenta la cobertura temática de acuerdo a los términos explorados del tesauro AAT, representado gráficamente de manera proporcional por el número de recursos digitales que se encontraron en Europeana. En este proceso se identificaron cuatro períodos predominantes con un elevado número de recursos digitales: “*Iron Age*”, con un total de (11.222) recursos digitales (24%) del total de recursos digitales extraídos; “*International Styles*” con un total de (7.337) recursos digitales (16%); “*Bronze age*” con un total de (5.833) recursos digitales (12%) y “*Abstract*” con (2.981) recursos digitales (7%). Estos resultados indican un alto porcentaje de recursos digitales cubiertos por estas temáticas de acuerdo a la clasificación definida por tesauro AAT. Sin embargo, en este gráfico también se identifica un gran número de términos representados por círculos más pequeños, que indica un número proporcionalmente bajo de recursos digitales identificados en Europeana para los términos del tesauro AAT asociados a estos círculos.

En la Tabla 4 se presenta un resumen de los resultados obtenidos a partir de una clasificación definida por el nivel de cobertura, en una escala de 0 recursos digitales (Ausencia de cobertura) a más de 10.000 recursos digitales (alta cobertura).

Tabla 4. Nivel de cobertura temática de los términos del tesauro AAT explorados en Europeana

Nivel de Cobertura Temática	Número de recursos digitales encontrados en Europeana	Número de Términos del Tesauro AAT explorados en Europeana	Porcentaje de cobertura del total de recursos digitales explorados
No hay cobertura	0	14	12%
Cobertura Baja	1-100	64	54%
Cobertura Media	100-1,000	34	28%
Cobertura Alta	1,000-10,000	7	6%

De acuerdo con la clasificación de la Tabla 4, y al concepto de cobertura definido en la Sección 4.2.1, se logró identificar un alto grupo de términos del tesauro AAT (64) que tenían una baja cobertura (54%), un pequeño grupo de términos (7) con un alto nivel de cobertura de recursos digitales en Europeana (6%) y 34 términos con un nivel de cobertura medio (28%). Por otro lado, un grupo de términos (14) no se encontraban asociados a ningún recurso digital de Europeana, lo cual nos indica que no disponían de cobertura temática.

A partir de los resultados de cobertura temática podemos observar en primera medida, que un gran conjunto de temáticas del tesauro AAT asociadas al área de conocimiento de “Estilos y Periodos”, no se encuentran cubiertas en su totalidad por Europeana. Sin embargo, por otro lado la variedad de resultados desplegados en el proceso de exploración de recursos digitales, dan a entender que Europeana no dispone de métodos de búsqueda de recursos digitales asociados a un área de conocimiento. Por lo tanto, para llevar a cabo este tipo de

comprobaciones fue necesario establecer una serie de estudios complementarios mediante un análisis de calidad de los metadatos explorados de acuerdo a criterios de evaluación asociados con la *completitud* (identificar el porcentaje de campos cuyo contenido es un valor no nulo) y la *exactitud* (identificar el nivel de precisión con que se describe un recurso digital). Una de las posibles razones que puede explicar esta variación de resultados se puede encontrar en la deficiente calidad de los metadatos definidos en recursos digitales explorados. Por tal motivo, estos resultados fueron base para llevar a cabo un proceso de análisis de calidad de metadatos con el propósito de: i) estudiar la *completitud* y *precisión de los metadatos* definidos sobre recursos digitales explorados (descrito en mayor detalle en la Sección 5.3), y ii) identificar criterios de búsqueda asociados con aquellos elementos de metadatos que presenten un mayor grado de *completitud*. Este último criterio se describe en la siguiente sección.

4.2.4 Definición de criterios de búsqueda

Los criterios de búsqueda basados en metadatos se han planteado en varios estudios dentro del campo de bibliotecas digitales (Stan Ruecker, Radzikowska, & Sinclair, 2011). Este tipo de estrategias facilita procesos de búsqueda de recursos digitales (Chirita, Gavriloiu, Ghita, Nejd, & Paiu, 2005; Hughes, 2005; Lytras & Sicilia, 2007; Moen, Stewart, & McClure, 1997; Ochoa & Duval, 2006; Park, 2005; Sánchez & Sicilia, 2005; Stawniak, 2012). Por tal motivo, para la definición de criterios de búsqueda visuales, se llevó a cabo un estudio preliminar de calidad de metadatos de los recursos digitales extraídos relacionados a la *completitud*. El concepto de *completitud* se encuentra asociado con el número de campos cuyo contenido es un valor no nulo (Moen, et al., 1997; Ochoa & Duval, 2006; Park, 2005). En el caso de los campos de varios valores, el campo se considera completo si existe al menos un valor registrado. De acuerdo con autores como Ochoa & Duval (2006) esta métrica presenta una primera aproximación para evaluar la completitud de un registro de metadatos.

A partir de este criterio de medición, se realizó un estudio inicial orientado al análisis de *completitud* asociados a un conjunto de metadatos clasificados en grupos *obligatorios* y *recomendados* de acuerdo al modelo de intercambio de datos de Europeana EDM (Figura 23) sobre el primer proceso de extracción de recursos digitales (555.000) realizado en Europeana. El propósito de este análisis es identificar qué metadatos presentan un mayor grado de completitud y adicional a ello, evaluar si son metadatos válidos para definirlos como criterios de búsquedas de recursos digitales a partir de las interfaces gráficas de visualización. La Figura 23 muestra los resultados de este análisis de *completitud*.

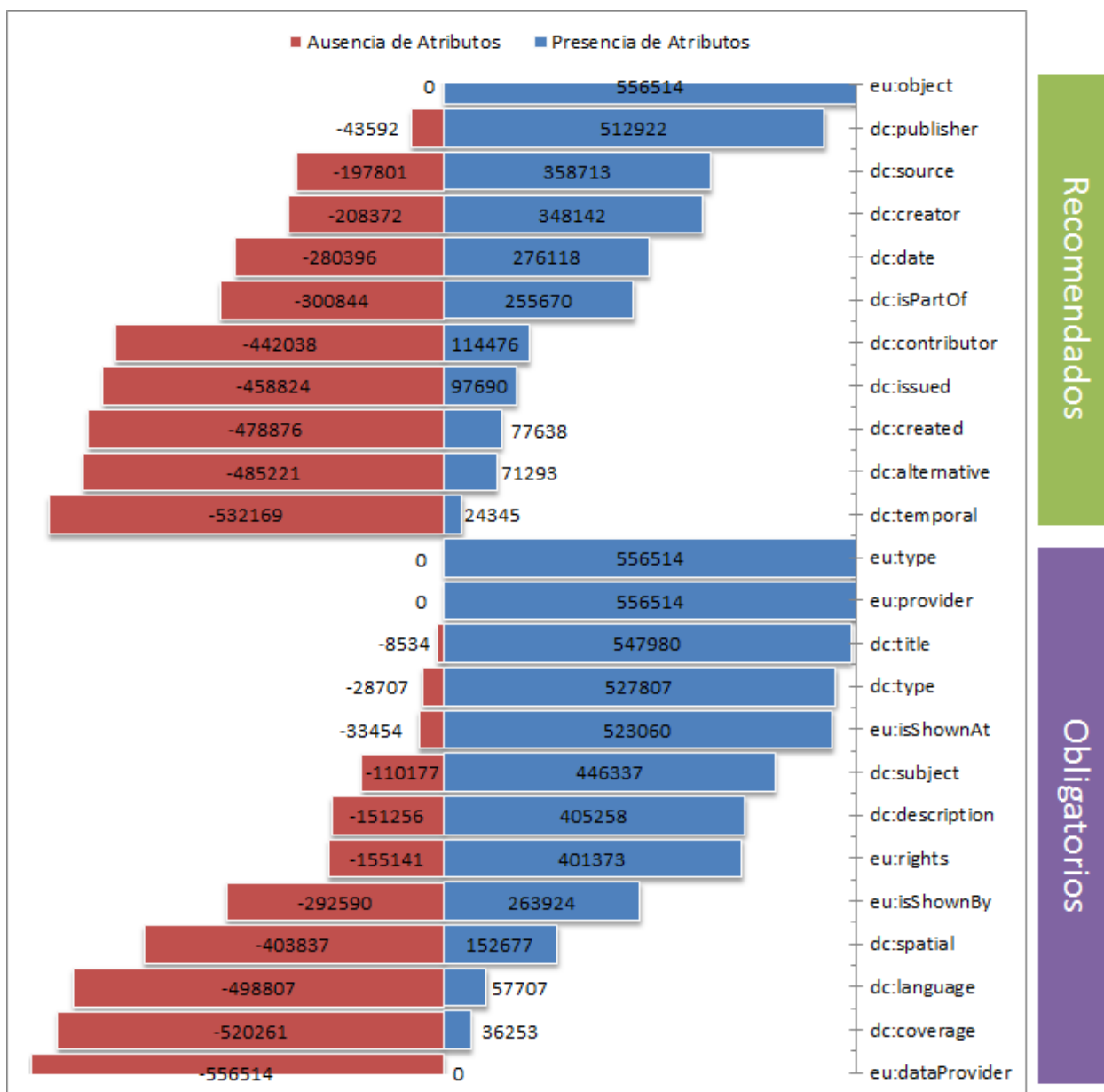


Figura 23. Análisis de completitud para el conjunto de metadatos obligatorios y recomendados

Uno de los resultados preliminares que arrojó este estudio, es el conjunto de metadatos con un porcentaje de *completitud* de 100%. La Tabla 5, presenta los resultados para el conjunto de metadatos que presentaron un mayor porcentaje de *completitud* en la descripción de recursos digitales.

Tabla 5. Listado de metadatos con mayor porcentaje de completitud

No	Nombre del Elemento	Registros Llenados	%
1	europa:provider	556,514	100.00%
2	europa:country	556,514	100.00%
3	europa:language	556,514	100.00%
4	europa:type	556,514	100.00%
5	dc:title	547,98	98.47%
6	dc:identifier	543,221	97.61%
7	dc:type	527,807	94.84%
8	europa:isShownAt	523,06	93.99%
9	dc:publisher	512,922	92.17%
10	dc:subject	446,337	80.20%

A partir de los resultados de la Figura 23 y la Tabla 5, los metadatos que presentaron un mayor porcentaje de completitud se encontraban directamente relacionados con elementos de metadatos definidos por Europeana. Este conjunto de metadatos son: *proveedor, país, idioma y tipo de recurso digital*. A partir de estos resultados, se tomará como referencia este conjunto de metadatos, para definir los criterios que se integrarán dentro de las estrategias de búsqueda visual implementados sobre los esquemas de representación de conocimiento. Este proceso se describe con mayor detalle en la siguiente sección.

4.3 Diseño de interfaces de navegación

Para esta fase de trabajo, en la que se llevó a cabo la implementación de las interfaces de navegación como estrategia para búsquedas visuales, se parte de varios criterios de diseño asociados al desarrollo de interfaces visuales y al uso de esquemas de representación de conocimiento a partir de taxonomías y tesauros. Dentro del diseño de interfaces, se tomaron en cuenta principios de Interacción Persona Ordenador para facilitar procesos de navegación e interacción con el usuario. Por lo tanto se definieron varias áreas de trabajo definidas por:

- *Área de exploración:* Se identificaron diversos métodos de exploración (*browsing*), los cuáles según Marchionini (1997) permiten definir estrategias de localización de información, y de acuerdo a (Y. Lin, et al., 2010) facilitan la visualización de grandes cantidades de información y percibir estructuras o relaciones jerárquicas. De acuerdo con Marchionini (2006), este criterio dentro de una interfaz con características ricas semánticamente, tiene el potencial de ayudar a los usuarios a la formulación de preguntas y llevar a cabo un proceso de exploración de acuerdo a una visión completa de las colecciones de recursos digitales disponibles.

- *Área de navegación:* Se pretende desplegar la información mediante mapas de categorías según la estructura taxonómica. Esta actividad facilita la navegación semántica de contenidos de acuerdo a un dominio de conocimiento (Card, et al., 1999) para facilitar a los usuarios la ubicación de ciertos términos.
- *Área de interacción:* Se determinaron estrategias de *zooming* para ampliar o disminuir el mapa de navegación y *panning* para recorrer toda la estructura de términos a consultar. A través de la estrategia de *zoom* se aplicaron técnicas de *foco* y contexto (J. Lamping, Rao, & Pirolli, 1995), las cuales permiten tener un efecto de enfocar el nodo que se está visualizando, permitiendo el desenfoque de otros nodos asociados a otros niveles de jerarquía.

A partir de estos criterios, partimos del uso de estrategias de búsqueda basadas en la exploración y soportadas por actividades de navegación a partir de técnicas de visualización gráfica, con el fin de explotar las capacidades de exploración de un usuario (Marchionini, 2006). Las estrategia de búsqueda exploratoria ha sido estudiada en varios escenarios de trabajo presentando buenos resultados (White, Kules, & Drucker, 2006; White & Roth, 2009). White (2006) resalta la importancia de estas estrategias de búsqueda, dado que presentan la capacidad para que los usuarios puedan consultar y navegar dentro de una estructura definida, fomentando de esta manera el aprendizaje y la investigación. Resultados que han sido soportados por estudios relacionados a nivel de exploración y navegación sobre estructuras taxonómicas (Uddin & Janecek, 2007; Z. Wang, et al., 2008), al igual que la importancia de las tareas de búsqueda basadas en el aprendizaje, como las estrategias más adecuadas para llevar a cabo actividades basadas en navegación (Marchionini, 2006).

A continuación se presentan las actividades relacionadas a esta fase que permitieron complementar el diseño de interfaces de búsqueda visual basados en técnicas de visualización.

4.3.1 Diseño de estructura jerárquica y proceso de transformación de datos

En la revisión de la literatura de la Sección 2.6.2 se encontraron una serie de principios para analizar el uso de taxonomías, con el fin de identificar formas de representación taxonómicas para mejorar los experimentos planteados. Estos principios estaban relacionados con: i) el diseño de un esquema de representación simple y granularidad de las categorías (Morante, 2003) con el fin de facilitar una estructura de navegación usable, ii) los niveles de navegación, con el objetivo de analizar la complejidad de la profundidad de una estructura y ruta de navegación (Bruno & Richmond, 2003), y iii) los niveles de exploración, con el fin de identificar la mejor estructura de navegación de acuerdo con la profundidad de la taxonomía (Z. Wang, et al., 2008).

Partiendo de estos principios, adicionalmente para el diseño de las interfaces visuales se tomaron en cuenta dos conceptos claves. La primera es la idea de interfaces enriquecidas visualmente, donde las representaciones individuales de cada elemento de una colección de

recursos digitales se complementan con herramientas emergentes de búsquedas (Stanley Ruecker, 2003). El segundo conjunto de principios que aplicamos son las ideas de diseño de interfaces de búsqueda a partir del uso tesauros (Shiri, Revie, & Chowdhury, 2002), dentro de las cuales resaltamos:

- La integración del diccionario de sinónimos y recuperación de documentos en pantalla.
- Formas flexibles para la selección de términos de consulta.
- Proporcionar listas jerárquicas y alfabéticas para apoyar diferentes estrategias.
- La descripción de un término y su localización dentro de una estructura jerárquica.

Para la definición de la estructura taxonómica se parte del uso de tesauros bien definidos. Esto presenta facilidades a nivel de: i) estructura navegacional, ii) jerarquía, y finalmente iii) por sus capacidades semánticas detalladas en la Sección 2.6. En la Figura 24, se presentan las dos ramas de conocimiento principales seleccionadas para el experimento (“*Styles and Periods by era*” y “*Styles and Periods by region*”) definidos en el Nivel 1 de clasificación. A continuación se presentan algunos términos del total de 118 seleccionados para llevar a cabo el proceso de exploración de recursos digitales, de acuerdo a la definición jerárquica presentada por el tesauro AAT.

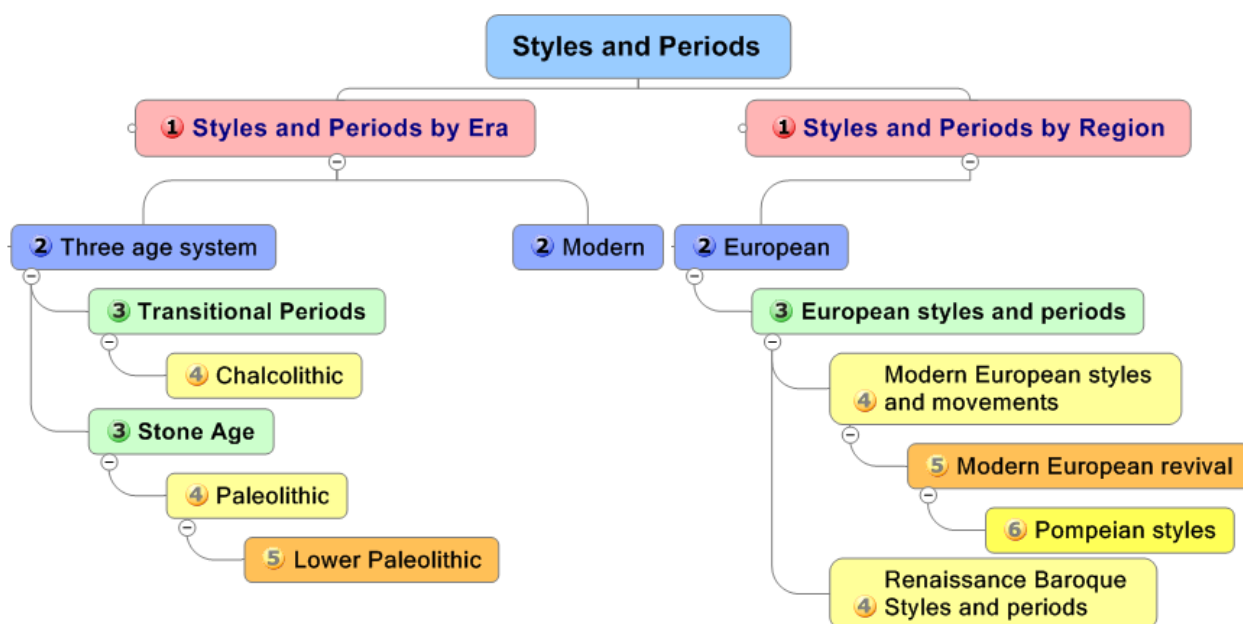


Figura 24. Niveles de profundidad de la estructura jerárquica definida por el tesauro AAT

Para crear una estructura jerárquica a partir de una estructura taxonómica se realizó un proceso de transformación de datos. Para ello se tomó como referencia el formato de representación de mensajes Java Script Object Notation JSON (Crockford, 2006); formato que facilita el intercambio de una estructura de datos a través de una arquitectura de tipo cliente - servidor mediante lenguajes de scripting como JavaScript, Python o Ruby. En la Figura 25 se presenta este proceso de transformación.

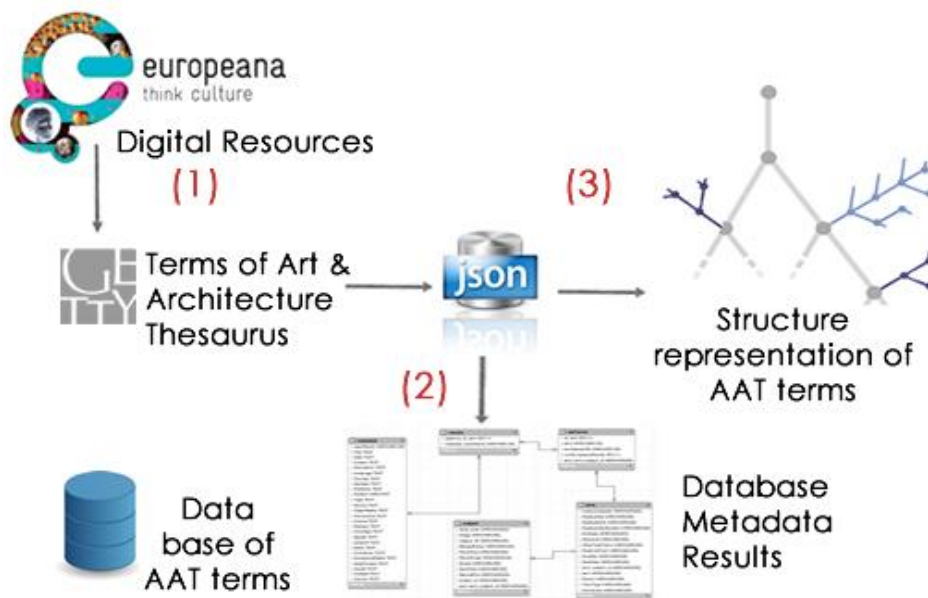


Figura 25. Proceso de análisis y transformación de datos

En la Figura 25 se presenta el proceso de transformación de los términos seleccionados del tesaurus para llegar a la definición de la estructura jerárquica de representación de datos a partir del formato JSON. Esta estructura de datos, fue la base para llevar a cabo un proceso de enlace a partir de las técnicas de visualización seleccionadas para llevar a cabo actividades de navegación y posterior vinculación con mecanismos de búsqueda de términos y recursos digitales. Este proceso se presenta con mayor detalle en la siguiente sección.

4.3.2 Selección de bibliotecas de funciones e implementación de técnicas de visualización sobre interfaces de navegación

Actualmente existe un gran número de proyectos que presentan herramientas y un conjunto de bibliotecas de funciones que aplican métodos y técnicas de visualización. Sin embargo existen una variada lista de factores y criterios que influyen sobre la selección de este tipo de herramientas y bibliotecas de funciones para llevar a cabo procesos de visualización, algunos de ellos analizados y mencionados por investigadores en el área (Chen, 2004; Fekete, 2004; Harger & Crossno, 2012; Hetzler & Turner, 2004; Shneiderman & Plaisant, 2006).

Para nuestro caso, se ha centrado en el uso de APIs con soporte para visualizar información sobre la Web, e independientes del uso de complementos o descargas de plugins adicionales para el despliegue de las visualizaciones. Por lo tanto, las interfaces de visualización se desarrollaron a partir de un API Open Source definida en el proyecto Infovis^[4], el cual de acuerdo a sus funcionalidades, presenta características de escalabilidad, facilidad de uso, alto rendimiento y reutilización de bibliotecas de funciones (Fekete, Hémerly, Baudel, & Wood, 2011). Otra de las ventajas al usar este tipo de bibliotecas de funciones, es el soporte que posee para trabajar sobre la mayoría navegadores en Internet, facilitando la independencia de

plugins adicionales para visualizar las interfaces de búsqueda, y mejorando el rendimiento a nivel computacional para llevar a cabo actividades de interacción sobre cada una de las acciones a realizar por parte del usuario. Otra de las características ofrecidas por esta API, es el soporte de mayoría de técnicas de visualización para la diagramación de estructuras taxonómicas basados en bibliotecas de funciones gráficas, las cuales disponen de aplicaciones de diferentes funciones computacionales a partir de la definición de bibliotecas de funciones que combina la diagramación de vectores a través de Vectores Gráficos Escalables por sus siglas en inglés SVG (Scalable Vector Graphics) y la combinación de tecnologías basadas en HTML5 (W3C, 2013) y componentes basados en cavas (Fulton & Fulton, 2013). Por lo tanto, este tipo de diagramación facilitó la construcción de las técnicas de visualización seleccionadas a partir de una estructura taxonómica de acuerdo a las necesidades del proyecto.

A partir del análisis presentado en la Sección 2.5 sobre técnicas de visualización, a continuación se presenta la implementación de ocho interfaces de navegación, visualmente clasificadas a partir de cuatro técnicas de visualización, para el despliegue del esquema de representación de conocimiento definido mediante formato JSON.

- 3) *Visualización tipo árbol*: Para el diseño de esta interfaz se aplicaron componentes adicionales que agregan criterios de selección a partir de su estructura de navegación, facilitando al usuario seleccionar el sentido de navegación por niveles de jerarquías: hacia abajo (Figura 26-a), arriba (Figura 26-b), izquierda (Figura 26-c) o derecha (Figura 26-d).

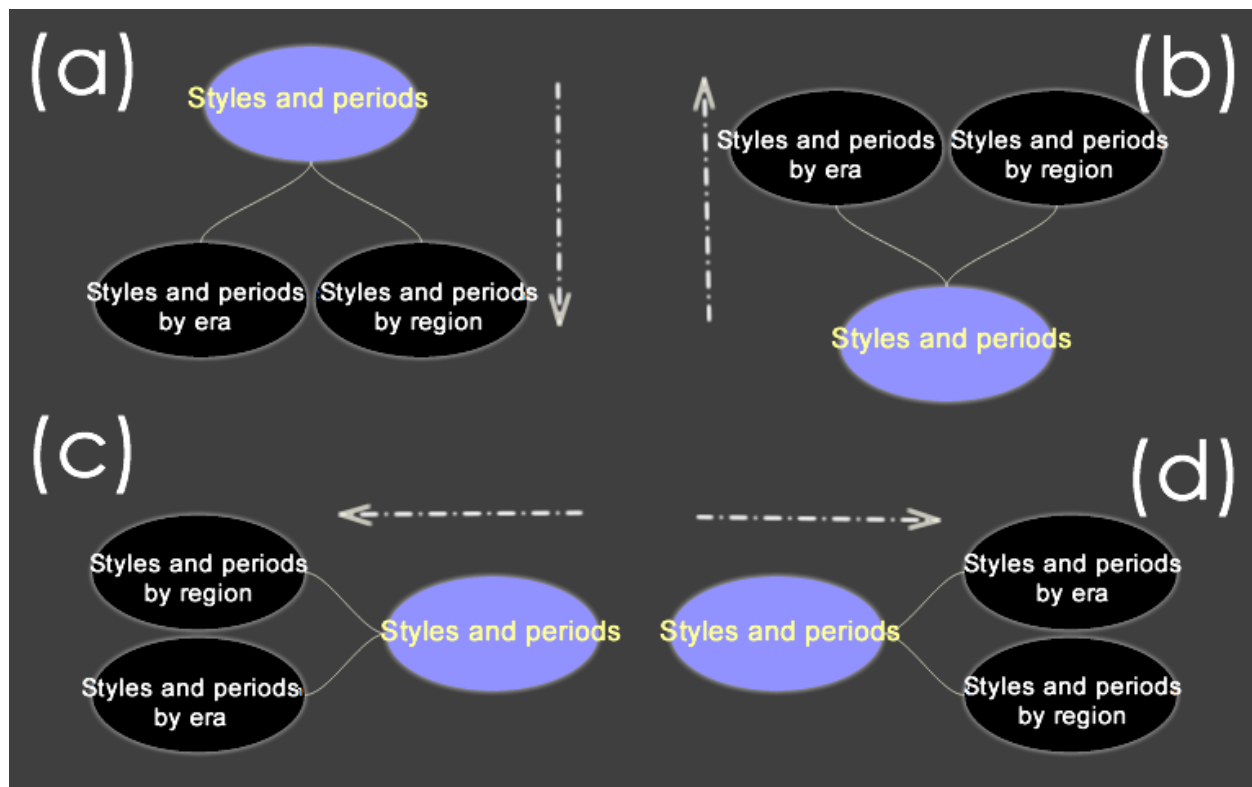


Figura 26. Interfaz de visualización tipo árbol

4) *Visualización tipo radial*: Para esta categoría, se diseñaron tres interfaces visuales a partir de tres estrategias de navegación y exploración diferentes:

- *Radial*: Técnica de visualización que se basa en la ubicación central de los nodos consultados, desplegando en la parte central de su estructura el nodo seleccionado por parte del usuario. Esto permite diferenciar cada nivel mediante la ubicación del nodo seleccionado en la parte central de la interfaz. La Figura 27 muestra un ejemplo del posicionamiento radial, presentando nodos en círculos concéntricos en función de su profundidad en el árbol. (Figura 27-a) nivel 1, (Figura 27-b) nivel 2, (Figura 27-c) nivel 3, Figura 27-d) nivel 4.

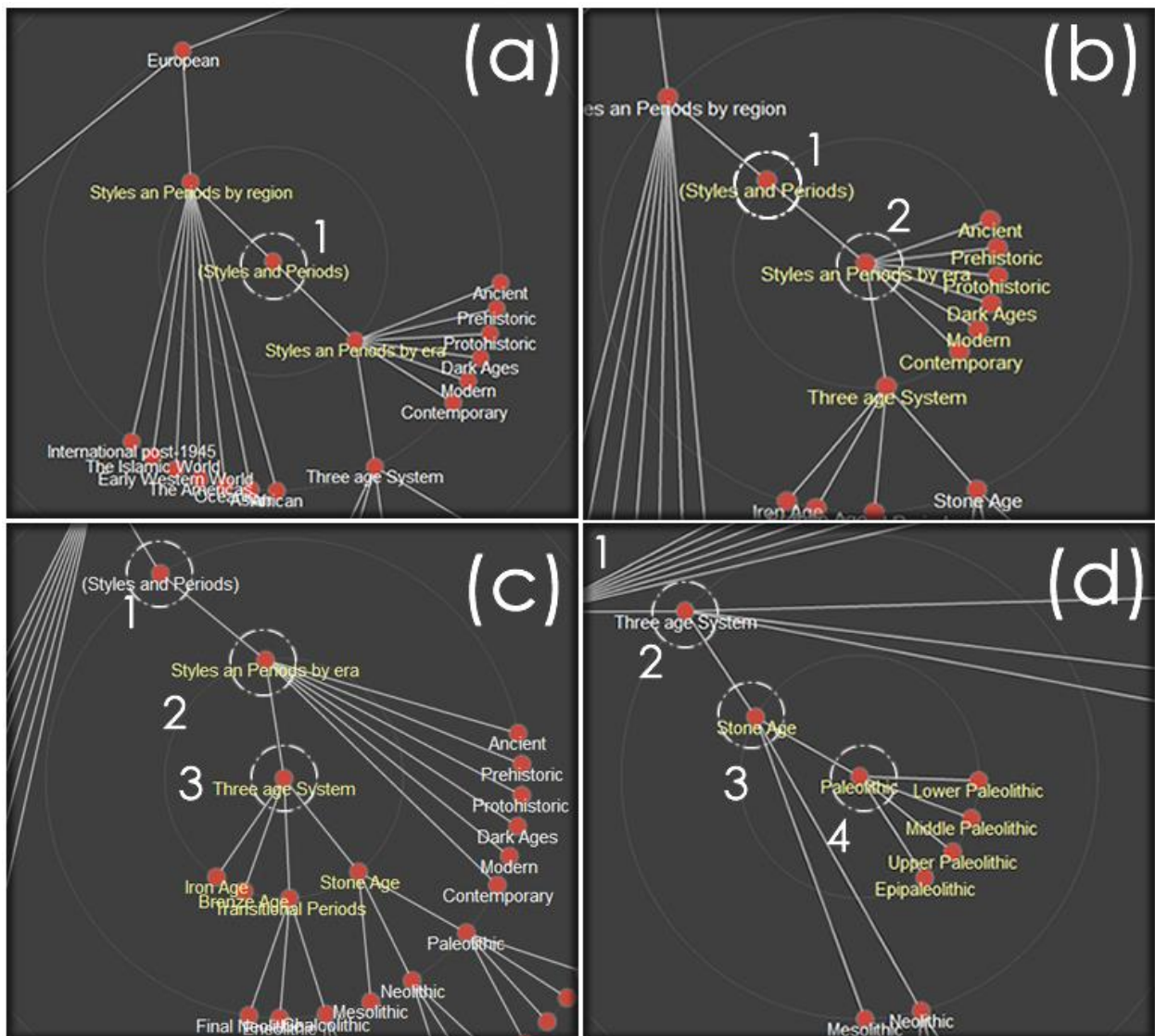


Figura 27. Interfaz de visualización “radial”

- *Radial-Search*: Visualización que permite la exploración de nodos en forma de radios, pero no se centraliza el nodo seleccionado, ni la profundidad de cada nivel explorado a diferencia de la interfaz “radial”. Otra diferencia respecto a la interfaz “radial” es el desarrollo de componentes adicionales relacionados a nivel gráfico con el tamaño de los nodos de acuerdo al número de recursos digitales asociados al término (a mayor número de recursos digitales, mayor es el tamaño del nodo seleccionado). De igual forma el nivel de exploración es acompañado a partir de la adaptación de un mecanismo de búsqueda mediante función de autocompletar (Figura 28-a). Esta estrategia complementaria ayuda el proceso de ubicación del término seleccionado dentro de la estructura de navegación (Figura 28-b).

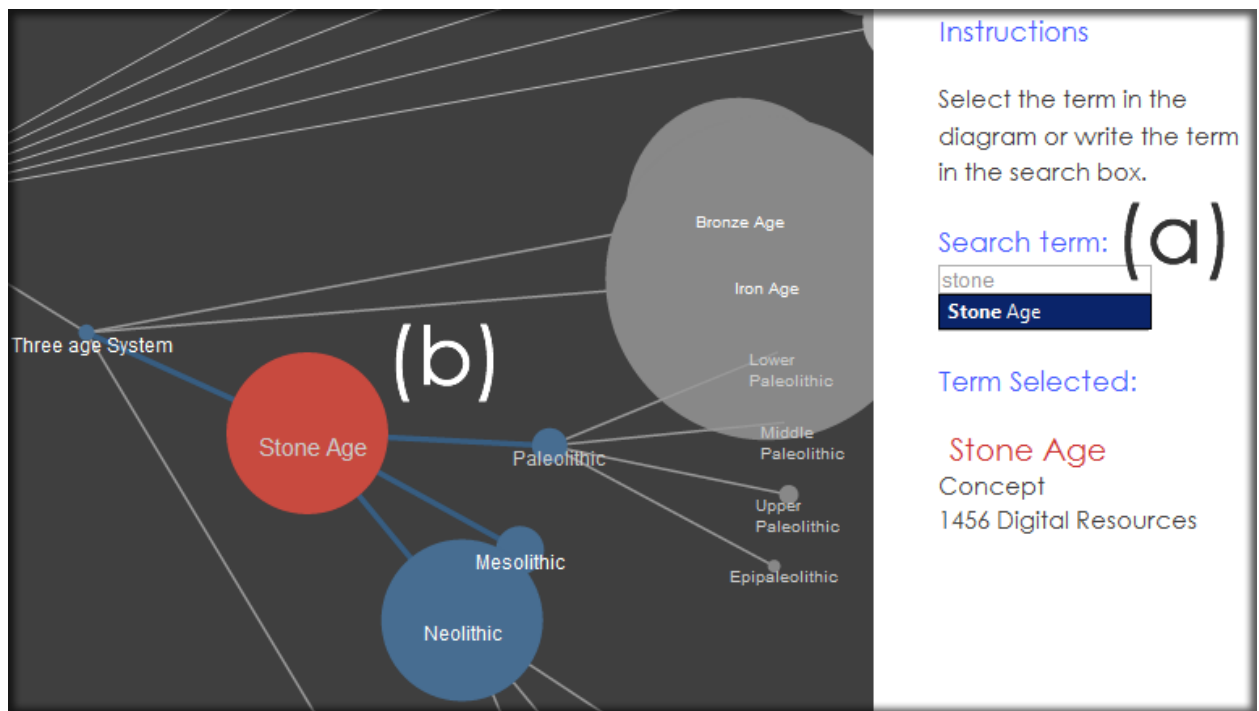


Figura 28. Interfaz de navegación “radial-search”

- *Relation*: Interfaz que permite desplegar un conjunto de términos asociados directamente a su nivel de jerarquía. Esto implica que las relaciones de niveles inferiores o superiores a la categoría seleccionada se ocultan automáticamente. Esta interfaz es una adaptación realizada a partir del proyecto Open Source Relation Browser^[5]. Adicionalmente es la única interfaz de navegación y búsqueda que trabaja mediante lenguaje ActionScript y bibliotecas de funciones multimediales soportadas por Flash. Además, la carga de datos y representación taxonómica se realiza mediante formatos de tipo XML. La Figura 29 (a) presenta la consulta de términos asociados a la temática “Three Age System”, donde se despliegan únicamente los términos relacionados a su nivel de categoría. Por su lado la Figura 29 (b), presenta los términos asociados a la sub-categoría “Stone Age”, ocultando los términos asociados a la categoría raíz “Three Age System” y desplegando los términos de la categoría “Stone age”.

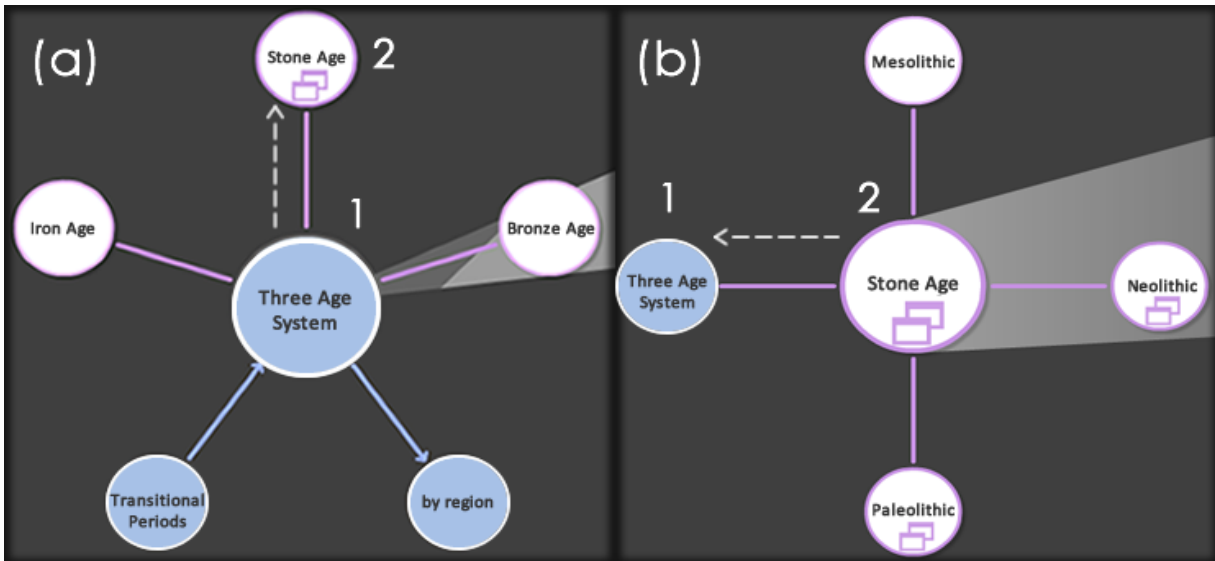


Figura 29. Interfaz de navegación “relation” (a) Nivel jerarquía superior (b) Nivel de jerarquía inferior

5) *Visualización tipo hiperbólico*: Para esta estrategia se trabajaron dos interfaces de visualización:

- *Sunburst*: Utiliza técnicas de foco y contexto, y estrategias de interacción mediante *zooming* y *panning* de manera elíptica. La Figura 30 presenta un ejemplo de esta interfaz, mediante el despliegue de información de radios concéntricos, y niveles de clasificación o profundidad (cuatro), a partir del nodo central hacia el disco exterior.

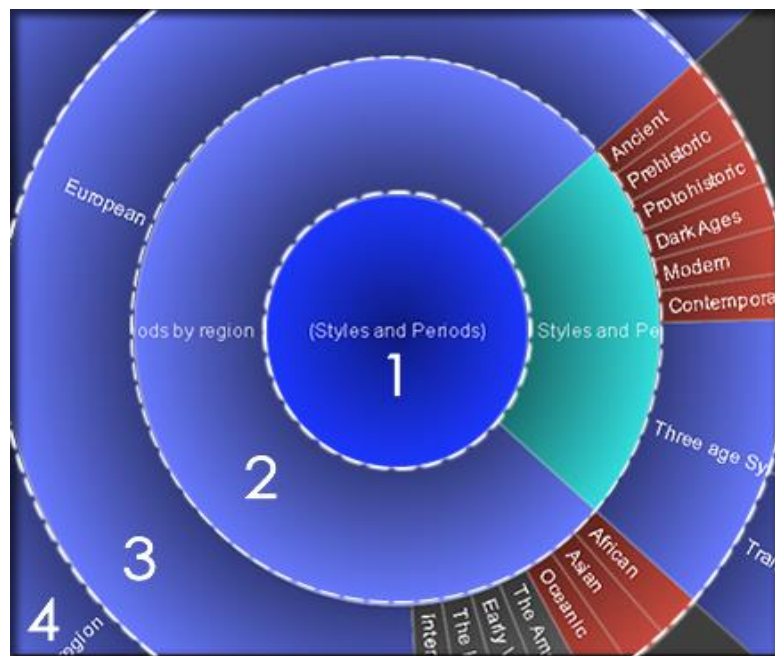


Figura 30. Interfaz de visualización “sunburst”

- **Hypertree:** Utiliza la técnica de foco y contexto representada a través del efecto ojo de pez (Furnas, 1986; Noik, 1993). Esta interfaz imita el efecto que tiene una cámara fotográfica al ampliar la vista del nodo que se está consultando, y ocultando el resto de las categorías representados en su estructura de navegación. A medida que se selecciona un término, se resalta junto con los que se encuentra directamente relacionados. La Figura 31-a presenta el nodo raíz, la figura 31-b, presenta el segundo nivel de profundidad. La figura 31-c, muestra el tercer nivel de profundidad, finalmente la figura 31-d despliega el 4 nivel de clasificación. A medida que se identifiquen mayores niveles de distancia al nodo seleccionado, se presenta el efecto de desenfoque visual. A medida que se explora, el nodo raíz se desenfoca completamente del mapa de navegación, facilitando al usuario centrar su percepción visual únicamente en el nodo seleccionado y los nodos directamente asociados al mismo.

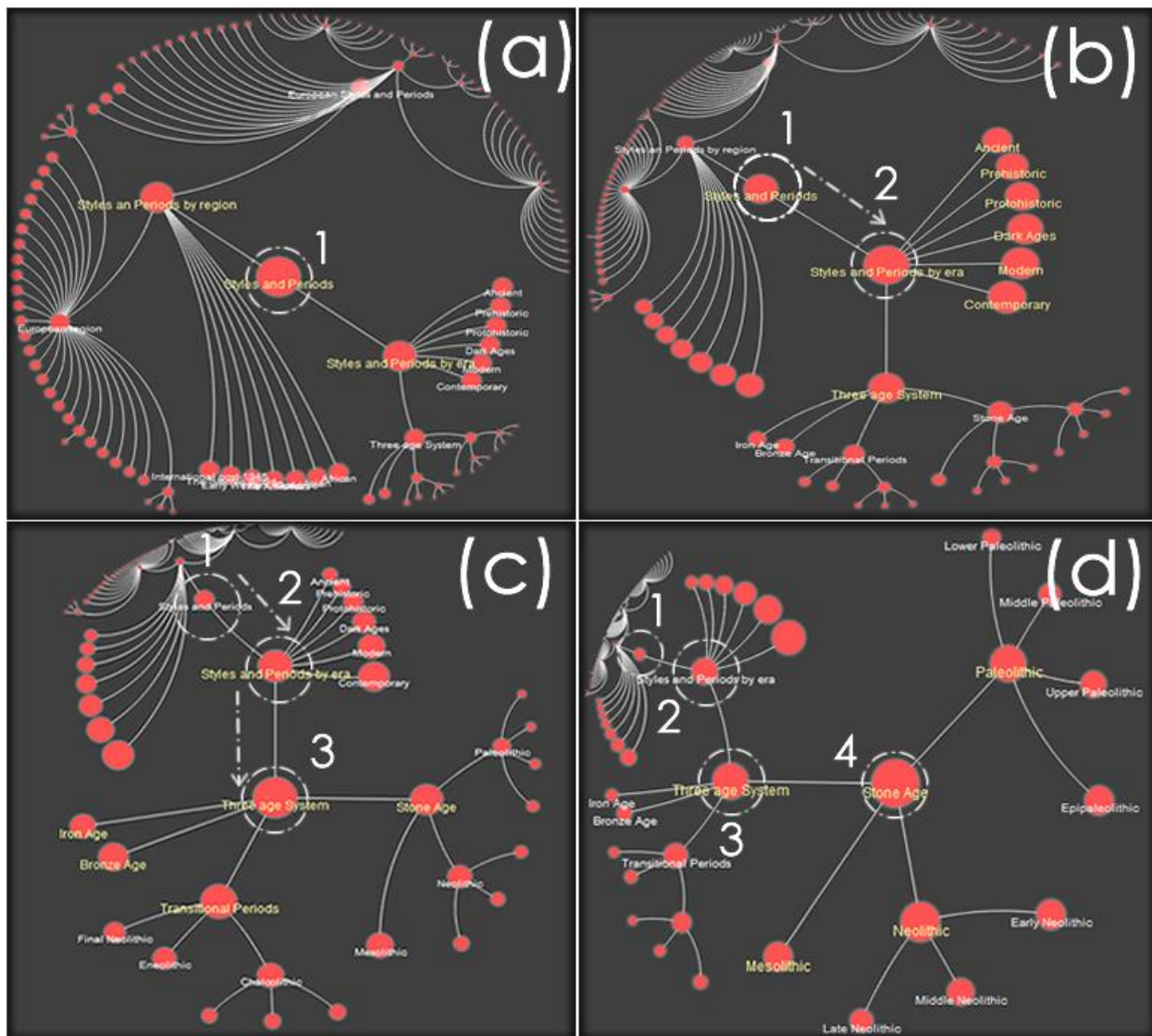


Figura 31. Interfaz de visualización "hypertree"

6) *Visualización de tipo Categoría:* Para esta estrategia se aplicaron 2 técnicas de visualización:

- *Treemaps:* Es una técnica de visualización que despliega una estructura jerárquica, categorizada sobre el total del espacio disponible en pantalla sobre una región rectangular (Shneiderman, 1991). La Figura 32, muestra un ejemplo de esta interfaz mediante el despliegue de información de manera jerárquica basado en la subdivisión recursiva del espacio de trabajo de manera rectangular. En la parte superior se muestra un rectángulo que representa una categoría principal “*three age system*” (nivel 1). Dentro de esta categoría encontramos una subcategoría “*Stone age*” (nivel 2), y este a su vez presenta dos subcategorías “*Paleolithic*” y “*Neolithic*” (nivel 3). De acuerdo al número de categorías que presente la rama de conocimiento, estos se sub-dividen en rectángulos más pequeños de manera proporcional de acuerdo al número de recursos digitales, y al número de términos asociados en la misma categoría.



Figura 32. Interfaz de visualización “treemaps”

- *Icicle:* Es una técnica de visualización que facilita la representación de términos mediante una agrupación jerárquica. Dentro del proceso de navegación se identifican qué objetos o nodos se unen o abandonan un grupo a medida que se selecciona sobre un nuevo nivel de jerarquía. La Figura 33 presenta un ejemplo de la implementación de esta técnica mediante el uso de bibliotecas de funciones que permiten avanzar a través de acciones definidas por el mouse a un nivel superior (clic

izquierdo) o retroceder de nivel (clic derecho) La Figura 33-a presenta el nodo raíz, la figura 33-b, presenta el segundo nivel de profundidad. La figura 33-c, muestra el tercer nivel de profundidad, finalmente la figura 33-d despliega el 4 nivel de clasificación. A diferencia de la interfaz *treemaps*, *icicle* realiza un proceso de navegación de izquierda a derecha o superior - inferior (dependiendo del sentido de navegación que se defina). Además de ello, cada categoría no se encuentra contenida dentro de una categoría principal, por el contrario cada nivel de clasificación va apareciendo a medida que el usuario selecciona un término.

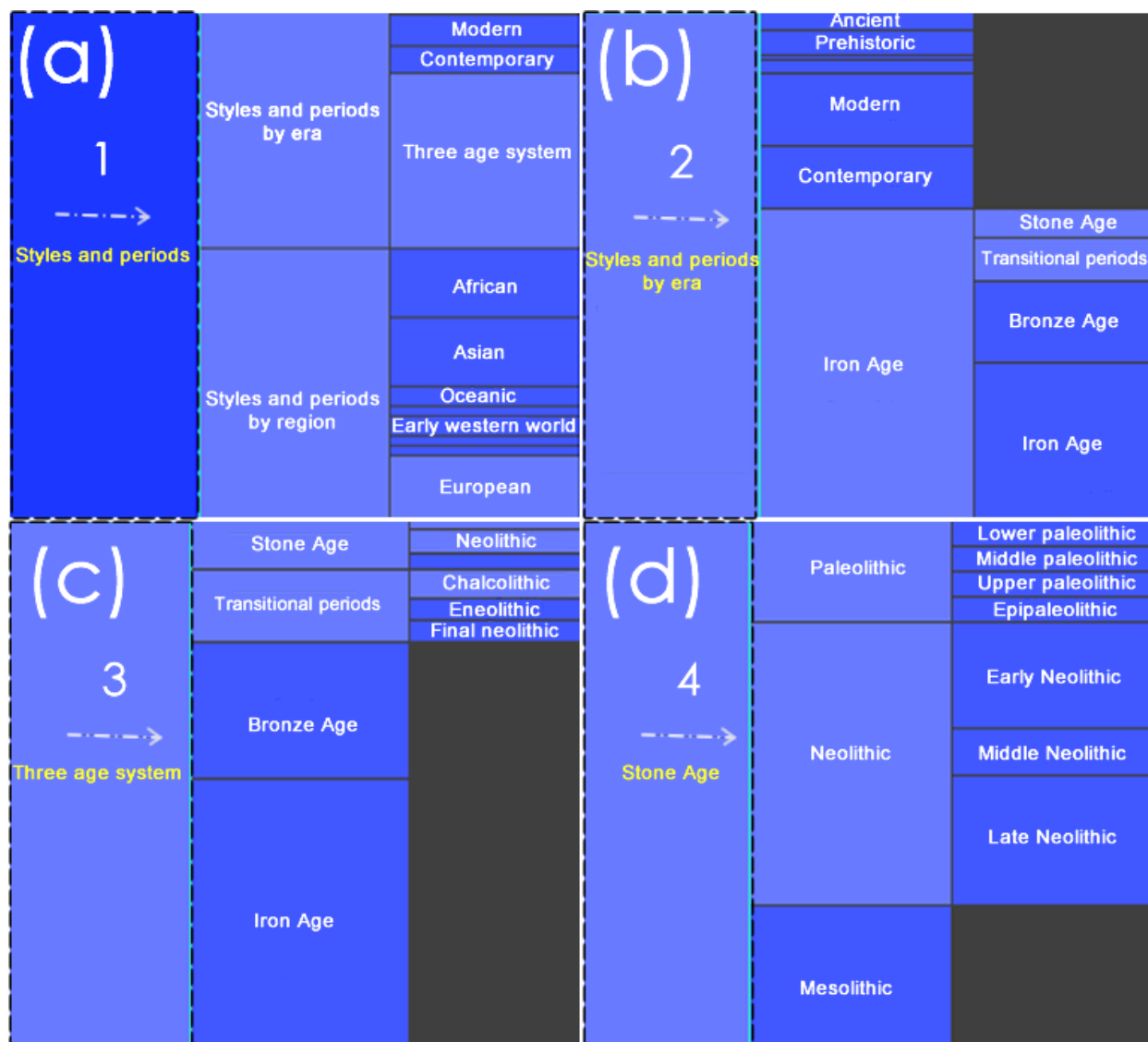


Figura 33. Interfaz de visualización "icicle"

Una vez diseñadas las interfaces visuales mediante técnicas de visualización, a continuación se presenta una propuesta de solución de búsqueda navegacional y exploratoria combinada con búsqueda textual y visual para la vinculación de recursos digitales a partir de criterios de búsquedas definidos por metadatos.

4.3.3 Diseño de un framework basado en interfaces de búsqueda

El framework planteado se apoya en cada uno de los resultados de estudios realizados en el Capítulo 6, para llevar a cabo la selección de interfaces aptas que faciliten procesos de búsqueda y acceso de recursos digitales a partir de:

- *Técnicas de visualización*: Integración de las mejores técnicas de visualización de acuerdo a resultados de usabilidad realizado en Capítulo 6.
- *Criterios de búsqueda visual*: Basados en metadatos a partir de áreas de conocimiento.
- *Alternativas de búsqueda textual*: Para la ubicación del término dentro de la estructura taxonómica.
- *Despliegue de recursos digitales mediante paginación*: Alternativa que carga los recursos digitales de acuerdo a criterios de selección de cada usuario.
- *Mecanismo de valoración de recursos digitales*: Para clasificar e identificar aquellos recursos más relevantes dentro de las primeras páginas de consulta de los recursos digitales desplegados.

Para la llevar a cabo su desarrollo se realizó la implementación de dos modelos de arquitectura de software, teniendo como propósito facilitar una estructura de aplicación genérica para la integración de otras técnicas de visualización y la vinculación posterior de más recursos digitales. Por un lado, se presenta una arquitectura Modelo Vista Controlador MVC encargada de establecer la estructura y base de la comunicación de la aplicación, y por otro lado se define una arquitectura encargada de comunicación con la parte visual, mediante el uso de bibliotecas de funciones basada en una arquitectura de tres capas.

A partir de un Modelo Vista Controlador MVC, se complementa la propuesta con una arquitectura de *tres capas*. Esta arquitectura además de definir un manejo de módulos independientes por niveles, mediante una comunicación bidireccional entre ellos, propone una baja dependencia en el modelo de comunicación, dado que los protocolos de comunicación que trabajan estas arquitecturas dejan en un segundo plano el uso tecnologías implementadas en cada una de las capas. De manera general, el modelo se resume en la Figura 34.

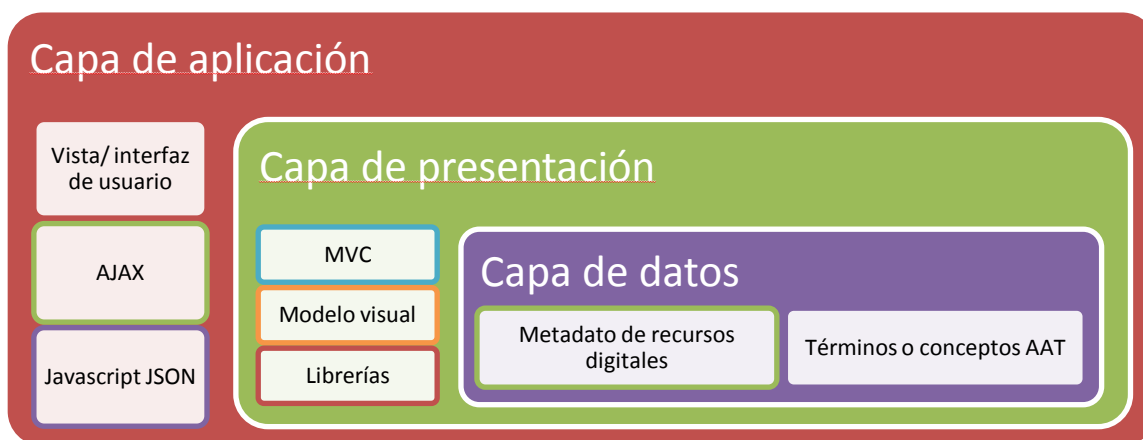


Figura 34. Arquitectura del modelo de tres capas para el área de visualización

- *Capa de almacenamiento de datos:* Presenta la colección de recursos digitales extraídos de Europeana, y los términos asociados al área de conocimiento del tesoro AAT.
- *Capa de presentación:* Presentan las bibliotecas de las técnicas de visualización aplicadas. Las solicitudes de consultas a nivel de usuarios relacionados con las palabras clave de un área del conocimiento, o de la selección de interfaces para mostrar la interfaz de visualización.
- *Capa de aplicación:* Presenta la petición de los usuarios y la selección de la interfaz visual para buscar recursos digitales de acuerdo a un área de conocimiento específica "Styles and periods".

Para llevar a cabo la integración de las interfaces, se han definido tres momentos o pasos de búsqueda para el acceso a recursos digitales a saber:

- *Primer momento:* Búsqueda de temática o área de conocimiento a partir de búsqueda textual o navegacional.
- *Segundo momento:* Selección de criterios de búsqueda basados en metadatos (país, idioma, tipo, proveedor) mediante el despliegue visual de manera proporcional del número de recursos digitales mediante el uso de diagramas de tipo pastel.
- *Tercer momento:* Localización de recursos digitales a partir de listado de búsqueda visual, mediante una previsualización de la descripción del recurso digital, previo a su selección.

En la Figura 35 se presenta un ejemplo del *primer momento de búsqueda* mediante la interfaz de búsqueda navegacional (radial).

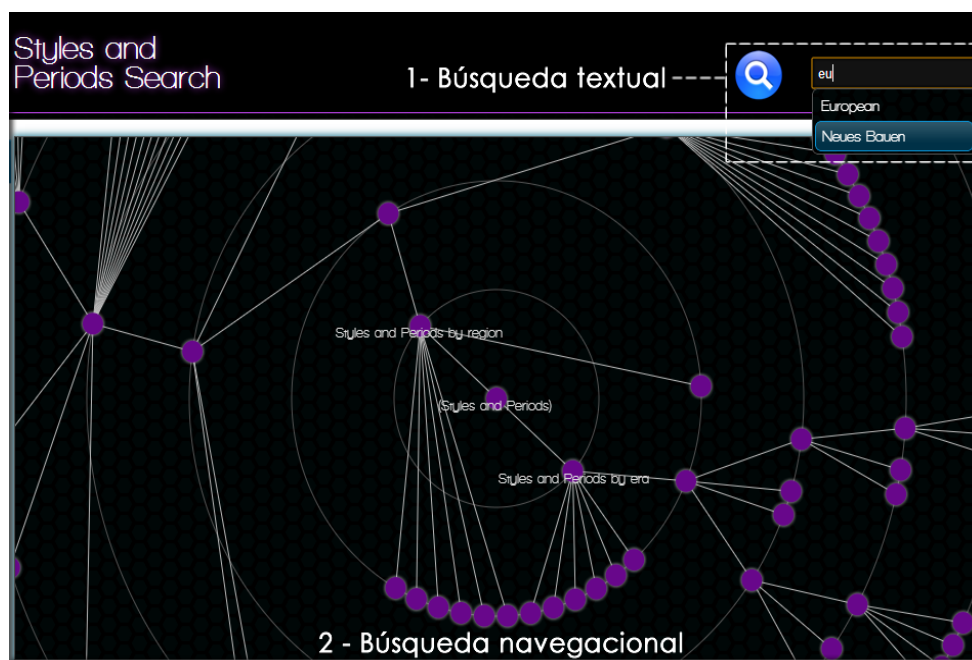


Figura 35. Integración de búsqueda textual y navegacional

En la Figura 35 puede observarse una muestra de la integración de búsqueda textual y navegacional para realizar un proceso de búsqueda a partir de un área temática previamente definido en un esquema de representación de conocimiento. En dicha Figura 35 se logra identificar:

- 1) Un área de búsqueda textual, el cual se encargada de autocompletar los términos de consulta del usuario de acuerdo al conjunto de 118 términos seleccionados del tesauro, y que se encuentran alojados en una base de datos.
- 2) Una vez identificado el término, este se localiza dentro de la estructura de búsqueda navegacional, definida a partir del esquema de representación de conocimiento.

Al localizar la temática de consulta, se presenta un *segundo momento*, donde el usuario puede seleccionar un criterio de búsqueda visual a partir de diagramas de tipo pastel, donde de manera proporcional se presenta el número de recursos digitales asociados a partir de sus metadatos (*Country, content provider, type, language*). En la Figura 36 se muestra un ejemplo de este momento de búsqueda, cuando el usuario selecciona la categoría “*styles and periods by region*”, desplegando en la parte inferior un conjunto de criterios de búsqueda a partir de los metadatos seleccionados en la Sección 4.2.3.

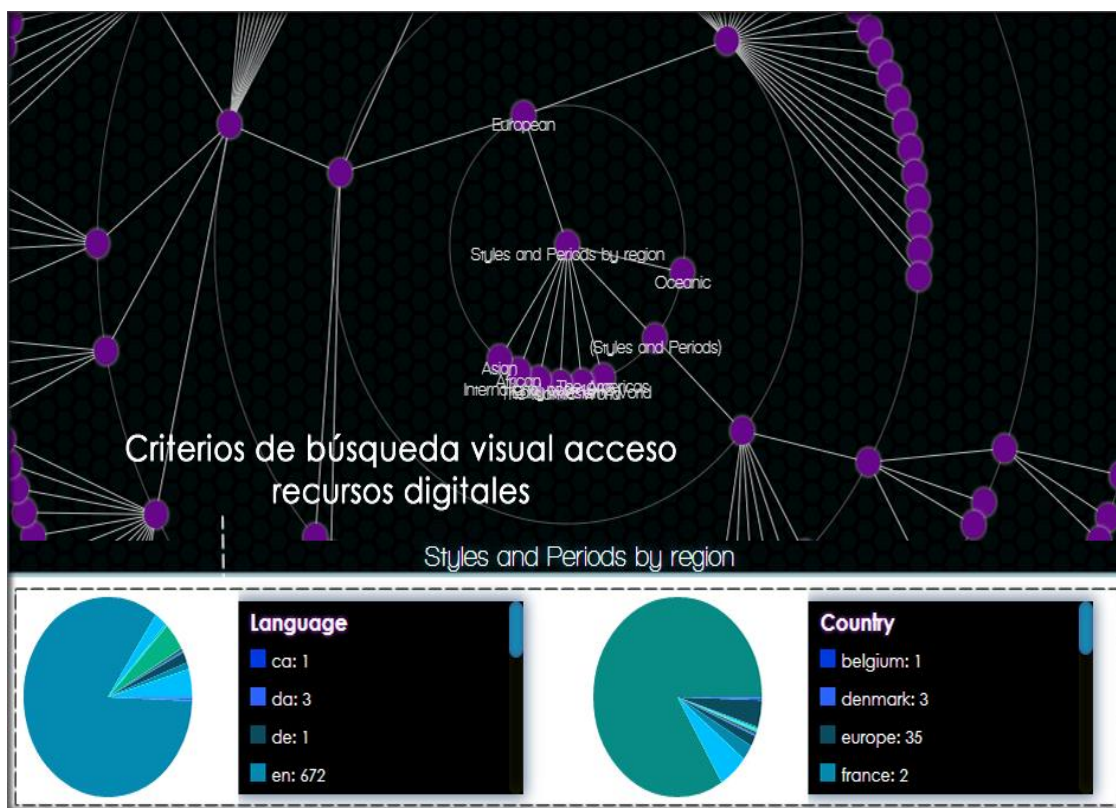


Figura 36. Criterios de búsqueda visual mediante diagramas de tipo pastel

El *tercer momento de búsqueda* se logra identificar cuando el usuario tiene la posibilidad de seleccionar dentro del diagrama pastel de acuerdo a criterios de búsqueda (*Country, content provider, type, language*). En la Figura 37 se presenta un ejemplo del despliegue de un listado de recursos digitales de la categoría “*styles and periods by region*”, de acuerdo al criterio de búsqueda por idioma en inglés.

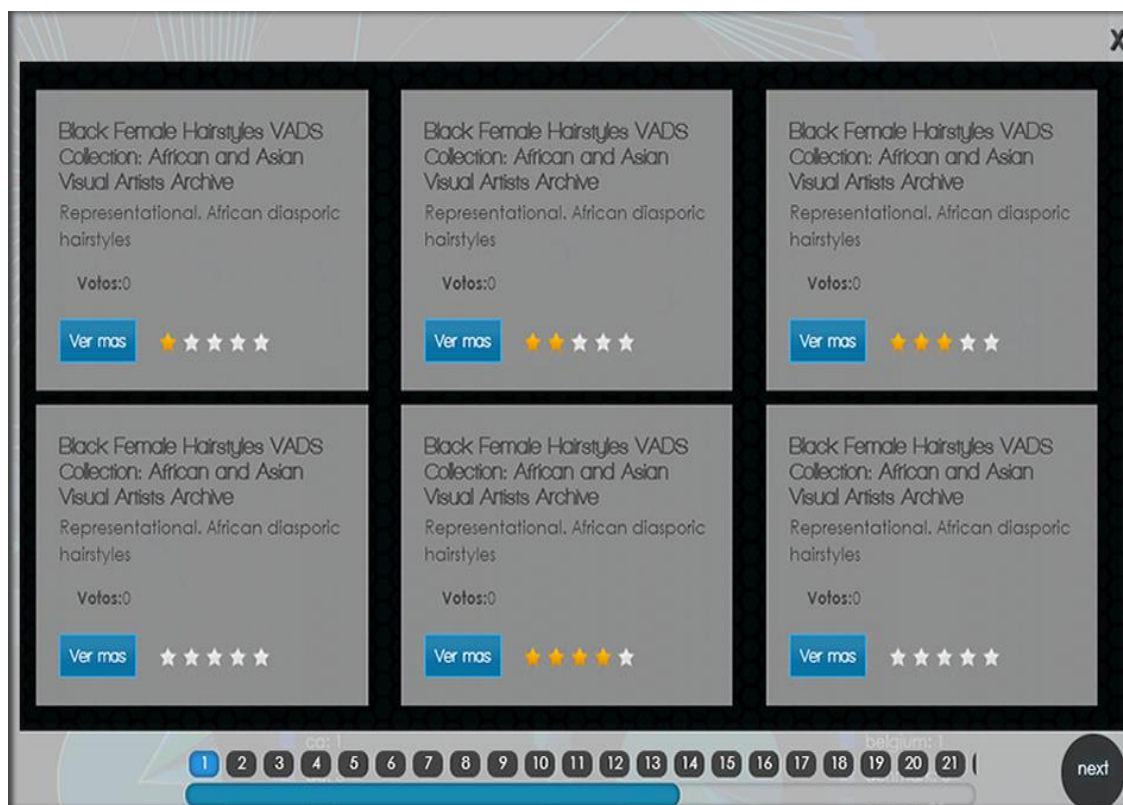


Figura 37. Despliegue de listado de recursos digitales

En la Figura 37 se muestra una carga de recursos digitales a partir de un listado de paginación, donde se despliega por cada página la descripción de 6 recursos digitales, y en la parte inferior se presenta un sistema de paginación para recorrerlos. Adicionalmente, se puede realizar una valoración del recurso digital con el propósito de que el mecanismo de despliegue ubique en las primeras listas de paginación la valoración realizada por cada recurso digital. Esto facilitaría la ubicación de recursos digitales relevantes de acuerdo a criterios de valoración realizada por usuarios que utilicen el repositorio digital.

Finalmente se presenta en la Figura 38 la descripción que se tiene de un recurso digital previo a su acceso. Se describe el título del recurso digital, el tipo de recurso digital, una breve descripción del contenido del mismo, y finalmente el enlace que permite conectar con la fuente del recurso digital para su consulta en el repositorio donde se encuentra alojado.



Figura 38. Descripción del recurso digital

A partir de este prototipo, se logró integrar métodos de búsqueda textual, junto con métodos de búsqueda navegacional a partir de esquemas de representación de conocimiento. Finalmente se logró integrar métodos de búsqueda visual, mediante diagramas de tipo pastel, el cual presenta de manera proporcional el número de recursos digitales asociados a cada criterio de búsqueda definidos de manera previa a partir de sus metadatos.

4.4 Resumen

En este capítulo se ha realizado una descripción y bosquejo de las estrategias que se llevaron a cabo para realizar el diseño e implementación de interfaces de búsqueda y navegación a partir de técnicas de visualización. En una primera sección se ha descrito la metodología que ha permitido desarrollar la propuesta de trabajo. Se ha presentado en primera instancia el planteamiento de un caso de estudio para el análisis y exploración de recursos digitales a partir de la biblioteca digital Europea con el fin de analizar su cobertura a partir de un conjunto de temáticas seleccionadas del tesoro AAT “*styles and periods*”. Debido a la variedad de resultados asociados a este estudio, fue necesario plantear un estudio complementario para llevar a cabo un análisis de metadatos de los recursos digitales explorados, con el fin de identificar qué metadatos podemos utilizar para definirlos dentro de los criterios de búsqueda gráficas que se desarrollarán dentro de las interfaces. Por otro lado se define un estudio

complementario que se abordará en la Sección 5.2, con el fin de evaluar la exactitud y completitud de los metadatos para identificar si los resultados obtenidos se deben a una mala calidad en la definición de los mismos.

Finalmente se presentó el diseño de las interfaces de búsqueda y navegación a partir de técnicas de visualización, partiendo por la definición del esquema de representación de conocimiento, seguido del proceso de transformación de los datos, el diseño de la estructura de navegación y niveles de clasificación, y posteriormente su integración dentro de las técnicas de visualización previamente estudiadas y analizadas en la Sección 2.5. En la parte final de este apartado se presentó el desarrollo de un prototipo de búsqueda visual, donde se integraron algunas técnicas de visualización a un framework que permite vincular búsquedas textuales y el despliegue de recursos digitales a partir de búsqueda visual mediante criterios de selección a partir de metadatos. Estas interfaces de visualización fueron seleccionadas de acuerdo a un estudio de usabilidad que de manera previa se llevó a cabo con un conjunto de 32 personas. Estudio y resultados que se presentarán en mayor detalle en la siguiente sección.

Pie de páginas:

- [1] Disponible en: <http://www.w3.org/RDF/>, (último acceso 5 Mayo 2014)
- [2] Disponible en: <http://www.w3.org/2004/02/skos/>, (último acceso 5 Mayo 2014)
- [3] Disponible en: <http://jericho.htmlparser.net/docs/index.html>, (último acceso 5 Mayo 2014)
- [4] Disponible en: <http://philogb.github.com/jit/>, (último acceso 5 Mayo 2014)
- [5] Disponible en: <http://moritz.stefaner.eu/projects/relation-browser/>, (último acceso 5 Mayo 2014)

Capítulo 5

Metodología y diseño experimental

El presente capítulo tiene como finalidad presentar el método de trabajo seleccionado para llevar a cabo el diseño experimental de las pruebas de usabilidad. Por lo tanto el propósito del capítulo se orienta a definir el diseño de las diferentes pruebas, que permitirán evaluar las interfaces de búsqueda para seleccionar aquellas que faciliten el acceso a una colección de recursos digitales a partir de criterios de búsquedas basadas en una temática o área de conocimiento, seguido de la selección de participantes y los test seleccionados para llevar a cabo las pruebas. Se presenta en un primer apartado la metodología que ha dado lugar para realizar el estudio de usabilidad. Para realizar este análisis, se planteó un caso de estudio mediante la participación de 32 usuarios. Se definieron dos etapas de análisis, a saber: percepción (Sección 5.3) e interacción (Sección 5.4). La finalidad de estas dos etapas de análisis tiene como propósito llevar a cabo la comprobación de resultados subjetivos (percepción) y resultados objetivos (interacción) mediante el planteamiento de un caso de estudio.

Capítulo 5: Metodología y diseño experimental

“El experimentador que no sabe lo que está buscando no comprenderá lo que encuentra”. (Claude Bernard)

5.1 Método de prueba y evaluación de interfaces de búsqueda navegacionales

La localización de información requiere de un conjunto de cualidades que dentro del área de Interacción Persona Ordenador podemos asociar con la *atención*, *comprensión*, *aprendizaje* y la *capacidad de recordar y reconocer la información*. Estas cualidades son esenciales para que un usuario identifique diversos elementos dentro de un proceso de búsqueda de información y así poder localizar y sintetizar los resultados obtenidos (Komlodi, Marchionini, & Soergel, 2007). Para identificar la utilidad de las interfaces de búsqueda navegacional implementadas a partir de técnicas de visualización con el objeto de evaluar qué interfaces son más útiles y eficientes, se llevó a cabo un estudio de usabilidad mediante el planteamiento de un caso de estudio. Para ello, se definieron dos criterios de evaluación orientados al análisis de actividades asociadas con i) *percepción visual* e ii) *interacción* sobre un grupo de participantes que interactuaron sobre cada una de las interfaces diseñadas.

La primera parte del estudio (*percepción*, Sección 5.3) analiza la percepción visual de los participantes, mediante el diseño de unos escenarios de trabajo definidos a partir de imágenes estáticas, los cuales representan a todas las estructuras de navegación diseñadas. El objetivo de este test es analizar tres capacidades de percepción visual para el uso de interfaces de búsqueda, a saber: i) *atención*, ii) *retención de información* y iii) *comprensión*. La segunda parte del estudio (*interacción*, Sección 5.4) analiza el uso de las interfaces a través del planteamiento de un caso de estudio guiado por un moderador, para llevar a cabo actividades de búsqueda de un conjunto de términos dentro de las estructuras de navegación definidas a partir de las interfaces.

A continuación se presentan los criterios de análisis que se determinaron para plantear las pruebas de usabilidad, asociados con: i) selección de participantes, 2) test de percepción visual y iii) test de interacción.

5.2 Selección de participantes y pruebas

De acuerdo con los objetivos del análisis y con las recomendaciones definidas para este tipo de estudios de usabilidad (Nielsen, 1994a, 1994b), 32 participantes fueron seleccionados para llevar a cabo las pruebas. Todos los participantes eran de mediana edad, con estudios superiores, buen nivel de inglés, y fluidez en el manejo de aplicaciones Web para la búsqueda de información. Dada la naturaleza del estudio, no era necesario que los usuarios tuvieran un

conocimiento específico asociado con el uso de estructuras taxonómicas, ni de estrategias de búsquedas avanzadas. En su lugar, se prestó especial atención al conocimiento que tenían acerca de métodos de búsqueda, que junto con la información mencionada anteriormente se recogieron en un cuestionario (Anexo B). La Tabla 6 describe el perfil de los participantes.

Tabla 6. Perfil de participantes

Perfil	n	(%)	País	n	(%)
Investigadores	9	28.125	España	19	59%
Estudiantes de Master	7	21.875	Colombia	3	9%
Estudiantes Universitarios	6	18.75	Perú	3	9%
Estudiantes de educación secundaria	10	31.25	Ecuador	3	9%
			México	2	6%
			Cuba	2	6%
Total	32	100	Total	32	100

Por otro lado, a la hora de diseñar las pruebas se contó con el asesoramiento de expertos en usabilidad, que aconsejaron y colaboraron en su diseño. Previo a la realización de las pruebas, el moderador (experto en usabilidad) realizaba una breve explicación de la prueba, indicando el propósito de la misma y el tipo de actividades que debía realizar cada participante. El moderador situaba a cada participante en el escenario preparado con el propósito de que no tuviera presión al momento de realizar la actividad indicada. Para llevar a cabo las pruebas, se utilizó un lugar físico de trabajo acondicionado para cada participante y recomendaciones básicas de adecuación del espacio definidas por Bevan (1994)(Bevan & Macleod, 1994) (Ej: mesa de trabajo, videgrabadora, equipo portátil, lugar aislado del ruido). En la Tabla 7 se definen las características técnicas y herramientas de software utilizadas para llevar a cabo la prueba.

Tabla 7. Especificaciones de equipo y software de captura

Equipo y Software	Características
Procesador	Intel R Core [™] i5 2430 M CPU 2.40 GHz
Tarjeta Gráfica	RADEON Graphics 3000, 512 MB Memoria dedicada 1280 MB
Memoria RAM	8 GB
Tamaño de Pantalla	15,6 pulgadas
Resolución	1366x768, 32 bits de colores
Sistema Operativo	Windows 7 Professional SP3
Tipo Sistema	64 bits

Equipo y Software	Características
Operativo	
Conectividad	WIFI 802.11 b/n y LAN 10/100 Mbps
Cámara video	Integrada en equipo, 3.0 Mpx, micrófono integrado
Camtasia Recorder 8	Software video capturas sesión de pruebas

Las pruebas se definieron en dos partes a saber: i) mediante la selección de atributos para evaluar la percepción visual de los participantes, previa a la interacción de las interfaces, y ii) mediante criterios de usabilidad para evaluar las interfaces diseñadas a partir de la interacción de los participantes.

5.3 Test de percepción

El término *percepción visual* se ha utilizado en la definición de la primera prueba de usabilidad, para referirnos a la capacidad que tiene un usuario para interpretar información codificada y representada a partir de una imagen visual (Cleveland, Diaconis, & McGill, 1982). La evaluación del impacto que se lleva a cabo a partir de este tipo de estudios, permite a los diseñadores realizar procesos de optimización de sus visualizaciones de cara a mejorar el diseño de cada presentación visual (Mackinlay, 1991; Mackinlay, Hanrahan, & Stolte, 2007).

De acuerdo a las capacidades que ofrece este tipo de estudios (Cleveland, et al., 1982; Mackinlay, et al., 2007), y el soporte de expertos en usabilidad, el diseño de la prueba de *percepción visual* tenía como propósito que el participante pudiera realizar una asociación de semejanza de las técnicas de visualización implementadas en las interfaces de búsqueda, a partir de la representación de un conjunto de imágenes alusivas a cada una de ellas. Bajo este enfoque, cada imagen (escenario) representaba una relación de semejanza de una estructura de navegación jerárquica definida por una técnica de visualización. Por lo tanto, estos escenarios fueron diseñados a partir de algunos factores definidos por estudios previos (Lidwell, Holden, & Butler, 2010; Nielsen, 1994b) y basados en criterios definidos por expertos de usabilidad establecidos por:

- *Facilidad de temáticas:* Cada escenario diseñado debía representar ejemplos sencillos, para lo cual se adaptaron una serie de ejemplos cuyas temáticas se encontraran asociadas con áreas de conocimiento como: *geografía, deportes, películas, bebidas, animales, frutas, hogar y comidas* (una imagen por cada técnica de visualización). La Figura 39 presenta un ejemplo de un escenario (*películas*) definido a partir de la técnica de visualización *treemaps*.
- *Estructura jerárquica:* La estructura jerárquica de cada ejemplo debía ser igual para todos los escenarios definidos (3 niveles de profundidad).

- *Número de términos:* Cada escenario debía desplegar de manera proporcional el mismo número de términos definidos dentro de cada una de las estructuras de representación taxonómica (entre 12 y 13 términos).
- *Asociación temática:* Cada término definido dentro del escenario o representación taxonómica, debía mantener una relación de asociación respecto a todo el escenario de manera global y por cada nivel de jerarquía establecido.



Figura 39. Ejemplo de escenario “películas” a partir de interfaz treemaps

Debido a las características específicas de este tipo de pruebas, se acondicionó un espacio físico adecuado para llevar a cabo las pruebas dentro de un entorno de trabajo apropiado (Reilly & Inkpen, 2007). De la misma forma se ajustaron parámetros de diseño, dado que los estudios de percepción visual también se ven afectados por características propias de los datos, incluyendo efectos de contraste de imágenes, como por ejemplo, el uso de colores de fondo con alta luminancia (Spence & Lewandowsky, 1991), el tamaño de las imágenes (J. Heer, Kong, & Agrawala, 2009) y los niveles de escala y proporciones de los datos (Cleveland, et al., 1982).

Uno de los métodos usados con mayor frecuencia para llevar a cabo pruebas de percepción visual se conoce con el nombre de “five-second test” (Boren & Ramey, 2000; Dumas, 1999; Nielsen, 1994b). Este método muestra al usuario en la prueba durante cinco segundos una aplicación de software o interfaz gráfica, y después de este lapso de tiempo un experto realiza

una serie de preguntas con el propósito de obtener las primeras impresiones asociadas al diseño. Para propósitos de nuestro estudio, se utilizó este método para evaluar cuatro aspectos (*atención, retención de información, comprensión e impresión subjetiva*):

- *Atención*: Si el usuario era capaz o no de identificar la temática que representaba la imagen desplegada.
- *Retención de información*: Cantidad de términos que el usuario recordaba de la imagen.
- *Comprensión*: Qué relaciones lograba identificar el usuario dentro del escenario de manera global y de acuerdo a los tres niveles de profundidad que representaba cada escenario.
- *Impresión subjetiva*: Después de ver todos los escenarios, se pidió a los participantes realizar una valoración de cada una de los escenarios.

Debido al número de escenarios (8 en total) y al número considerable de términos por escenario (12 y 13 términos), se ha realizado una adaptación de esta prueba para llevarla a cabo en un intervalo de tiempo más largo (diez segundos). La Figura 41 presenta un ejemplo del ambiente de trabajo planteado para la selección de cada escenario.

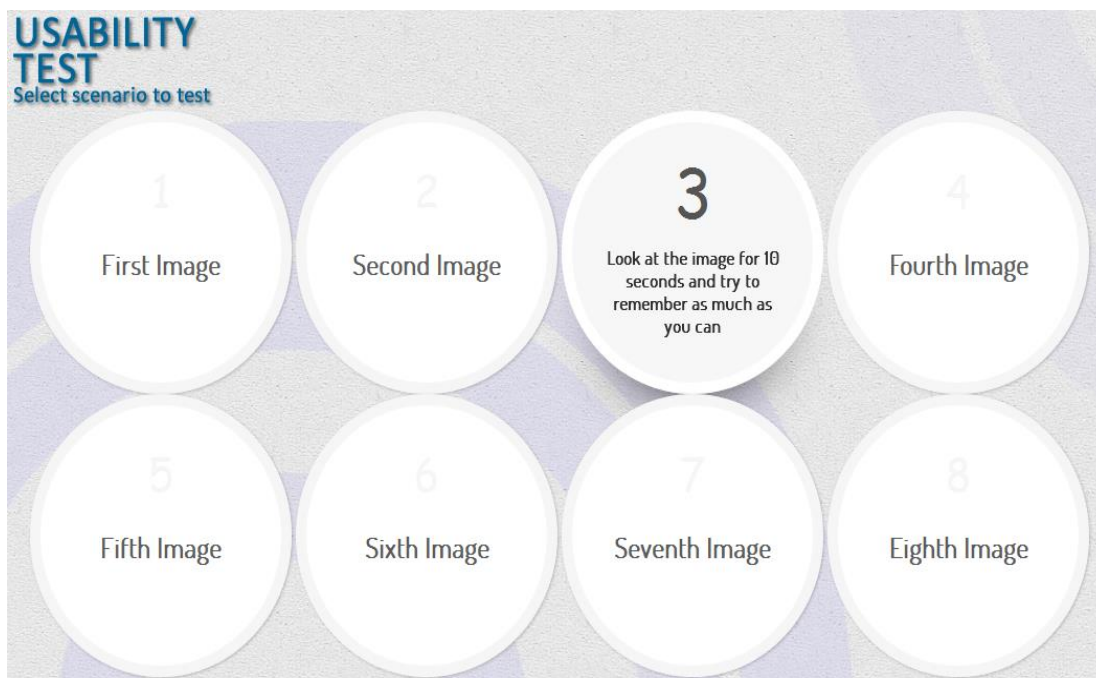


Figura 40. Interfaz de trabajo para realizar prueba de usabilidad de percepción visual

En este ejemplo (Figura 40) se pedía a cada participante seleccionar una opción que representaba un escenario de la prueba. Una vez transcurrido diez segundos, un moderador procedía a realizar una serie de preguntas relacionadas con el escenario desplegado con el propósito de evaluar los cuatro aspectos de usabilidad mencionados anteriormente (*atención, retención, comprensión e impresión subjetiva*). El proceso se detalla, junto con los resultados a modo de ejemplo, en la Sección 5.2.1. A continuación en la Tabla 8, se presenta el instrumento que permitía registrar las respuestas de los participantes.

Tabla 8. Instrumento para registrar interacción de percepción de usuarios

Identificador:	Nombre:
“TEN SECOND” TEST	

1. Relation

Temática	
Profundidad	Nivel 1
	Nivel 2
	Nivel 3
Recuerdo	Aciertos:
	Fallos:

2. Hypertree

Temática	
Profundidad	Nivel 1
	Nivel 2
	Nivel 3
Recuerdo	Aciertos:
	Fallos:

3. Hypertree

Temática	
Profundidad	Nivel 1
	Nivel 2
	Nivel 3
Recuerdo	Aciertos:
	Fallos:

4. Radial-search

Temática	
Profundidad	Nivel 1
	Nivel 2
	Nivel 3
Recuerdo	Aciertos:
	Fallos:

5. Radial

Temática	
Profundidad	Nivel 1
	Nivel 2
	Nivel 3
Recuerdo	Aciertos:
	Fallos:

6. Sunburst

Temática	
Profundidad	Nivel 1
	Nivel 2
	Nivel 3
Recuerdo	Aciertos:
	Fallos:

7. Treemaps

Temática	
Profundidad	Nivel 1
	Nivel 2
	Nivel 3
Recuerdo	Aciertos:
	Fallos:

8. Tree

Temática	
Profundidad	Nivel 1
	Nivel 2
	Nivel 3
Recuerdo	Aciertos:
	Fallos:

Finalmente, otro de los propósitos que perseguía este estudio era familiarizar al participante mediante la asociación visual de la estructura taxonómica, con el esquema jerárquico representado por cada interfaz como paso previo a la interacción real con cada una de las interfaces, para llevar a cabo procesos de búsqueda. Este último proceso se enmarca en una

segunda parte del estudio (test de interacción, Sección 5.4). De la misma forma, la prueba se enfocaba a evaluar cada actividad por parte de un moderador, quien era la persona experta y responsable del desarrollo de las pruebas. El moderador realizaba una serie de preguntas relacionadas con cada imagen que se presentaba, para llevar a cabo la evaluación de percepción e impresión subjetiva por parte del usuario; por ejemplo: indicar los términos que mayor recordaba de la imagen, ¿qué relación encontraba de los términos que recordaba, de acuerdo a cada nivel de profundidad que presentaba la imagen?, ¿cuál era la temática que de manera general describía toda la imagen?

5.4 Test de interacción

Después de realizar el test de percepción, se llevó a cabo un caso de estudio para realizar el proceso de interacción con las interfaces. El caso de estudio consistía en que cada usuario debía tomar el rol de un estudiante para identificar si la aplicación que se les presentaba, facilitaba la ubicación de una serie de términos y adicional a ello, si los resultados arrojados por el proceso de búsqueda podría determinar una buena cobertura de cada término indicado, de acuerdo al número de recursos digitales desplegados en el proceso de consulta. Por lo tanto cada participante debía localizar un conjunto de términos ubicados en diferentes niveles de jerarquía dentro de una interfaz, y realizar una serie de actividades orientadas a: i) localizar el término solicitado dentro de la estructura taxonómica, ii) identificar el número de recursos digitales asociados a cada término solicitado, iii) determinar qué tan cubierto se encontraba el término de acuerdo al número de recursos digitales (cobertura temática), y iv) evaluar la interfaz de acuerdo a criterios de usabilidad asociados con *facilidad de uso*, *navegación* y *estética*. Para realizar esta actividad se definieron dos fases de interacción (fase de entrenamiento y fase de pruebas).

5.4.1 Fase de entrenamiento

En esta fase, el participante seleccionaba una interfaz, y el moderador solicitaba buscar una serie de términos asociados a la rama de conocimiento de "*styles and periods by era*". Una vez localizado el término, se le pedía identificar el número de recursos digitales asociados. Una vez concluida esta actividad, el moderador planteaba una serie de preguntas asociadas a la cobertura de la temática consultada y el número de recursos, para determinar si el usuario era capaz de reconocer la información que la interfaz estaba desplegando.

Para evitar el sesgo que pudiera ocasionar la complejidad taxonómica, se consideró que cada término a consultar estuviera en niveles básicos de profundidad (niveles 2 y 3), con el propósito de ubicarlos en áreas claramente visibles dentro de la estructura de navegación. Al final de la prueba el moderador llevó a cabo una sesión de evaluación de los ejercicios realizados, mencionando los errores asociados a las respuestas dadas por cada participante y haciendo énfasis en los efectos de la ubicación de términos de acuerdo al concepto de cobertura

temática. A continuación (en la Tabla 9), se presenta el instrumento utilizado para registrar los resultados de la prueba.

Tabla 9. Instrumento para registrar fase de entrenamiento de interacción de usuarios

Identificador:	Nombre:
TEST DE USABILIDAD	

Para el siguiente test, no se necesita un conocimiento previo de las temáticas que se están presentando, simplemente están asociadas para medir el nivel de interacción del usuario con una serie de interfaces.

CASO DE ESTUDIO:

Estas ayudando a un estudiante de secundaria a preparar material de estudio para temas relacionados con el área de arte, arquitectura y periodos clásicos. Has localizado una página que trata el tema y tiene unos mapas conceptuales que muestran los recursos contenidos en la misma, quieres hacerte una idea de si reúne lo necesario como para recomendársela al estudiante. No debes acceder a ningún recurso concreto, simplemente decimos tu impresión sobre la cobertura, teniendo en cuenta que, como es evidente, no todos los recursos serían útiles.

PRUEBA ABIERTA (PARTICIPANTE)

1. Interfaz de búsqueda seleccionada:	Radial	Relation	Tree	Hypertree
	Icicle	Sunburst	Radial-Search	Treemaps

2. El alumno quiere hacer un trabajo sobre estilos de arte en las “tres edades” (Three Age System) en inglés, localice el término:

- **Tiempo:** _____
- **Errores:** _____

3. Basándose en lo que observa, indique si el alumno va a tener suficiente información para tratar el tema de las “tres edades” (Three Age System)

EDAD	IMPROBABLE	PROBABLE	SEGURO
Bronce (Bronce Age)			
Hierro (Iron Age)			
Era de Piedra (Stone Age)			

- **Tiempo:** _____
- **Errores:** _____

4. De los recursos identificados en las “sistema de tres edades” responda:

- ¿Cuál crees que tiene más recursos? _____
- ¿Cuál crees que está mejor clasificado? _____

5. De las tres edades, han solicitado que profundice en la “**edad de piedra**” dividiendo correctamente cada una de sus etapas. Indique si el alumno encontrará información suficiente para tratar los periodos de la tabla y reconocer sus subperiodos.

EDAD	IMPROBABLE	PROBABLE	SEGURO
Paleolithic			
Mesolithic			
Neolithic			

- **Tiempo:** _____
- **Errores:** _____

6. De los tres periodos tratados:

- ¿Cuál crees que tiene más recursos? _____
- ¿Cuál crees que está mejor clasificado? _____

5.4.2 Fase de prueba

Una vez concluida la fase de entrenamiento, se llevó a cabo una segunda prueba, donde el moderador asignó una interfaz diferente a la interfaz seleccionada por el participante durante la fase de entrenamiento. Se pidió a los participantes realizar el proceso de búsqueda de otro conjunto de términos asociados a la rama de conocimiento “*styles and periods by region*” y realizar la misma actividad de la fase de entrenamiento, pero para esta ocasión, cada término solicitado se encontraba ubicado en diferentes niveles de profundidad y de clasificación. La dificultad respecto a la fase de entrenamiento se determinó en la estructura taxonómica, entre 3 y 6 niveles de profundidad de la jerarquía. A continuación (en la Tabla 10), se presenta el instrumento utilizado para registrar los resultados obtenidos en esta fase de prueba.

Tabla 10. Instrumento para registrar fase de prueba de interacción de usuarios

TEST DE USABILIDAD

1. Interfaz de búsqueda asignada:

Hypertree	Radial	Relation	Tree
Treemaps	Icicle	Sunburst	R-Search

2. El alumno necesita realizar un trabajo sobre “**estilos y movimientos modernos europeos**” (**modern european styles and movements**) en inglés, localice el término:

- **Tiempo:** _____
- **Errores:** _____

3. Explore cada recurso, y de acuerdo a lo que observa, indique si el alumno tendrá suficiente información para abordar la temática de “**estilos modernos del renacimiento europeo**”:

EDAD	IMPROBABLE	PROBABLE	SEGURO
Arte Deco (Art Deco)			
Biedermeier			
Modernista (Modernist)			
Neo-Romántico (Neo-Romantic)			

• Tiempo: _____

• Errores: _____

4. De los términos asociados a la temática “**estilos y movimientos modernos europeos**”:

• ¿Cuál crees que tiene más recursos? _____

• ¿Cuál crees que se encuentra mejor clasificado? _____

5. Dentro de la categoría “**estilos y movimientos modernos europeos**” (**Modern european styles and periods**), han solicitado profundizar sobre “**estilos arquitectónicos modernos europeos**” (**modern european architecture styles and movements**). Indique si el alumno encontrará información suficiente para tratar esta temática de acuerdo a los términos de la siguiente tabla:

EDAD	IMPROBABLE	PROBABLE	SEGURO
Bellas artes (Beaux-Arts)			
Estilo Internacional (International Styles)			
Movimiento Moderno (Modern Movement)			
Bellas artes (Beaux-Arts)			

• Tiempo: _____

• Errores: _____

6. De los términos relacionados:

• ¿Cuál crees que tiene más recursos? _____

• ¿Cuál crees que está mejor clasificado? _____

Preguntas Generales:

1. ¿Considera que este tipo de interfaces son útiles para que un estudiante consulte el número de recursos y material didáctico para sus actividades académicas?

Sí _____ No _____ ¿Por qué? _____

2. ¿Considera que este tipo de interfaces son útiles para que un profesor o experto, les permita clasificar las temáticas de clase de acuerdo a un área en particular?

Sí _____ No _____ ¿Por qué? _____

3. De las interfaces vistas, puntúe del 1 (peor) al 5 (mejor):

INTERFAZ: _____	1	2	3	4	5
Navegabilidad					
Estética					
Facilidad para ver el número de recursos					
Facilidad para identificar la estructura de la temática					

4. De las dos interfaces probadas, globalmente, ¿cuál elegiría?:

La técnica empleada para llevar a cabo esta prueba fue “*thinking-aloud*” (Boren & Ramey, 2000) introducidas inicialmente por (Lewis, 1982), y validadas por otros autores como Nielsen & Hackos (1994). Esta técnica consiste en que cada usuario debe interactuar con el producto y participe pensado en alto, describiendo sus acciones, sentimientos y cualquier cosa que se le pase por la cabeza mientras va realizando el test (L. Castells, 2007). Por lo tanto se llevó a cabo un planteamiento de preguntas mientras cada participante interactuaba con la interfaz para localizar el término indicado, y así poder identificar más criterios de evaluación desde el punto de vista cualitativo para complementar el análisis de resultados obtenidos. Como es habitual en este tipo de estudios, toda la prueba se registró en vídeos capturando la interfaz de búsqueda y del usuario. Estos videos luego fueron revisados por parte del moderador y experto en usabilidad con el fin de extraer la mayor cantidad de información posible, con el propósito de que fuera de soporte para verificar los resultados obtenidos en la pruebas realizadas (ej. la atención dentro de una parte específica de la interfaz, o la expresión facial al interactuar con la interfaz).

Finalmente, se realizó un pre-análisis de los datos recogidos para preparar un grupo de discusión donde se incluyó la participación de seis usuarios que interactuaron con las interfaces. El propósito de este análisis era llevar a cabo una retroalimentación para profundizar más en los resultados, realizando preguntas asociadas con: i) el motivo por el cual realizaron algunas acciones dentro de la interacción, ii) las sugerencias que ellos indicaban para mejorar el proceso de exploración de recursos digitales, y finalmente iii) problemas dentro del proceso de exploración de las interfaces. La Figura 41 presenta un ejemplo de un momento de la prueba de interacción llevado a cabo a través de la interface *radial*. En la parte inferior derecha se muestra la captura de video del rostro de cada participante; en esta oportunidad se solicitaba al participante realizar la búsqueda del término “*three age system*” ubicado en el tercer nivel de profundidad dentro de la estructura taxonómica. La interfaz radial, muestra por defecto en la parte central del radio la raíz de la estructura taxonómica. Por lo tanto cada participante debía partir del término raíz “*styles and Periods*” (1), seguir hacia la siguiente rama “*styles and periods by era*” (2), y finalmente llegar al término solicitado “*three age system*” (3).

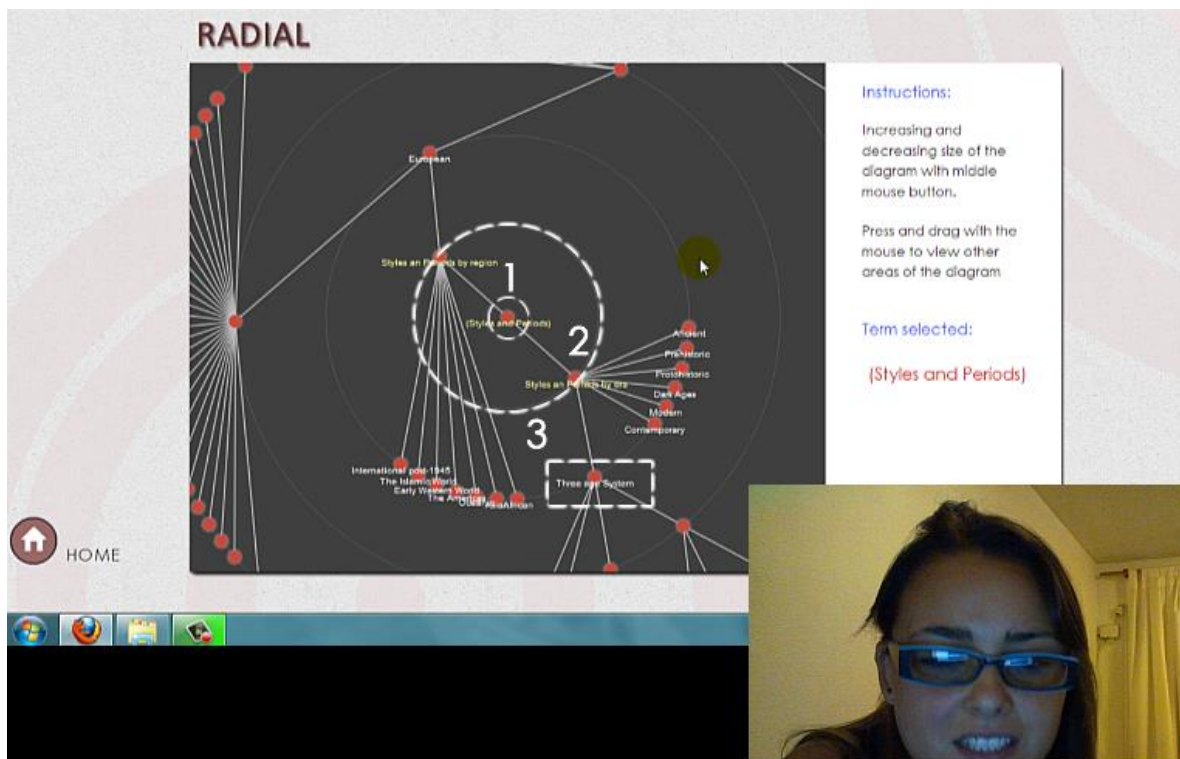


Figura 41. Momento de interacción de un usuario en prueba de usabilidad

Todo el proceso de la prueba es capturado mediante video y audio, tanto a nivel interno, grabando sobre el escritorio de trabajo; como externo, grabando imágenes y sonidos a través de la cámara y el micrófono integrados al portátil. De la misma forma la aplicación de captura almacena un registro donde se indica el tiempo que cada participante ha tomado para realizar la actividad indicada por el moderador.

En la etapa final de esta prueba, se plantearon una serie de preguntas sobre la impresión subjetiva y la experiencia adquirida por los usuarios. Estas preguntas fueron diseñadas para evaluar tres aspectos de usabilidad: *estética*, *navegabilidad* y *facilidad de uso*, y dos aspectos de la utilidad: la *cobertura temática* (número de recursos digitales asociados a un término) y el nivel de *clasificación* de los recursos consultados. En paralelo, se llevó a cabo un estudio de la *precisión* (si el usuario lograba localizar el término solicitado o no), a partir del número de éxitos o fracasos dentro del proceso de localización de términos en el transcurso de toda la prueba.

Posterior a la realización de estas pruebas de *percepción visual* y de *interacción*, se llevó a cabo un análisis de los resultados finales obtenidos, y de ello vamos a tratar en el siguiente capítulo.

5.5 Resumen

En el presente capítulo se presentó el proceso utilizado mediante la descripción de las diferentes fases de trabajo definidas, para llevar a cabo el desarrollo de las pruebas de usabilidad. En un primer apartado se logró definir las fases que dieron lugar a la realización del experimento, dividiendo así el experimento en dos pruebas diferentes a saber: (*percepción e interacción*). La primera prueba orientada de manera subjetiva, fue indispensable para medir el nivel de satisfacción de cada participante al hacer uso de cada una de las interfaces. La segunda, orientada de manera objetiva, con el propósito de obtener resultados más cercanos al uso y utilidad de cada una de las interfaces mediante el planteamiento de un caso de estudio. Ambas pruebas se lograron complementar para llevar a cabo un análisis de resultados, los cuáles se abordarán en el siguiente capítulo.

Capítulo 6

Discusión y análisis de resultados

Este capítulo tiene como propósito presentar los resultados de análisis y evaluación de estudios que permitieron proponer una solución orientada a facilitar a creadores de repositorios digitales, el diseño de las interfaces de búsqueda implementadas a partir de técnicas de visualización. Por lo tanto se presenta en un primer apartado los resultados de análisis y evaluación de los estudios de usabilidad descritos en el Capítulo 5. Estos resultados permitieron identificar de manera preliminar una serie de interfaces útiles para facilitar el proceso de búsqueda de recursos digitales, a partir de una visualización previa del número de recursos digitales.

En un segundo apartado, se presenta de manera paralela los resultados de análisis de calidad de metadatos (mencionados en la Sección 4.2.4.), con el propósito de evaluar la completitud y exactitud de los metadatos extraídos de la biblioteca digital que se tomó como caso de estudio (Europeana). Se pretende por un lado, determinar una serie de propuestas que permitan mejorar la clasificación y búsqueda de recursos digitales en un repositorio digital a partir de una temática o área de conocimiento. Por otro lado, identificar criterios de búsqueda a partir de los resultados de completitud de metadatos con el fin de vincularlas dentro de las interfaces de búsqueda visual.

Capítulo 6: Discusión y análisis de resultados

*“El que refrena su boca y su lengua se libra de muchas angustias”
(Proverbios 21:23)*

6.1 Análisis de los resultados del estudio de usabilidad

En esta sección se presentan los resultados del estudio de usabilidad en dos etapas. En la primera etapa (Sección 6.1.1) se presentan los resultados de las respuestas relacionadas con las impresiones subjetivas y capacidades de percepción de los usuarios. En la segunda etapa se analizan los aspectos de interacción mediante un análisis estadístico realizado sobre los atributos que tienen una mayor influencia en el grado de usabilidad y utilidad de acuerdo con la percepción de cada usuario identificado a partir de la cobertura temática (Sección 6.1.2), una evaluación objetiva sobre los resultados de precisión (Sección 6.1.4), y finalmente los análisis correspondientes a las pruebas de la hipótesis planteada en nuestro estudio (Sección 6.1.5).

6.1.1 Percepción e impresiones subjetivas

Como resultado del análisis de percepción (Sección 5.3), al final de este proceso se pidió a cada participante realizar sus impresiones subjetivas sobre las interfaces. La Figura 42 muestra en promedio los resultados obtenidos, ordenados de 1 (interfaz mejor valorada) a 8 (interfaz peor valorada).

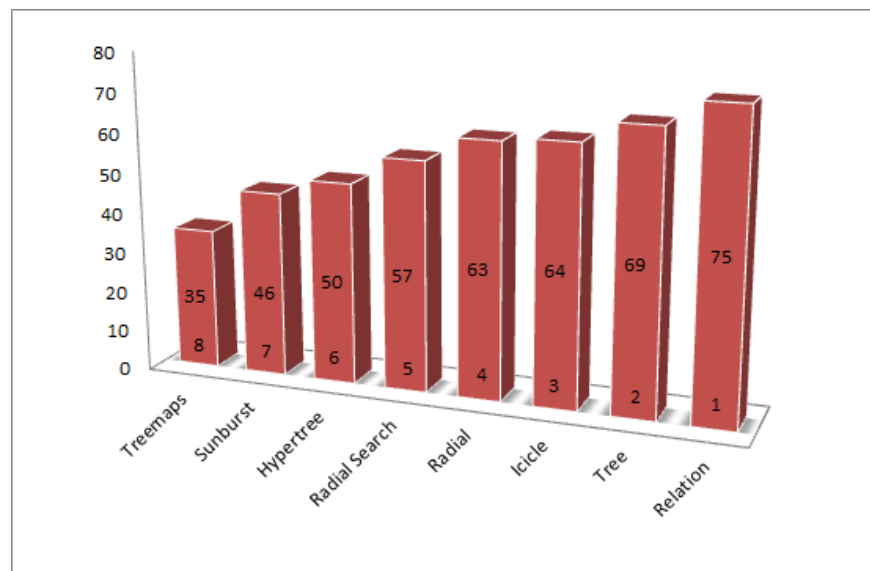


Figura 42. Clasificación de las interfaces de acuerdo a la percepción de los participantes (1 = más alto rango, 8 = más bajo ranking)

Durante la fase de evaluación de la usabilidad de percepción, las interfaces que recibieron las puntuaciones más altas fueron *relation* (1), *tree* (2) e *icicle* (3). Sin embargo, esta clasificación sólo muestra las impresiones subjetivas de los participantes de una manera generalizada. En la Tabla 11 se presentan los resultados obtenidos de acuerdo a los criterios de estudio asociados con la capacidad de percepción *atención*, *retención* y *comprensión*, definidos en la Sección 5.1.2.

Tabla 11. Resultados estadísticos asociados a la capacidad de percepción

Interfaz	Estadística Descriptiva	Atención	Retención	Entendimiento		
				Nivel 1	Nivel 2	Nivel 3
Tree	Media	100	47.3	100	68.8	93.8
	D.E.	0	10.8	0	0.47093	0.24593
Treemaps	Media	100	57.8	100	100	0.6
	D.E.	0	15	0	0	0.50402
Sunburst	Media	100	57.5	100	87.5	81.3
	D.E.	0	15.2	0	0.33601	0.39656
Radial	Media	100	49.3	100	87.5	84.4
	D.E.	0	13.3	0	0.33601	0.3689
Radial-Search	Media	100	63.3	100	81.3	40.6
	D.E.	0	12.9	0	0.39656	0.49899
Icicle	Media	100	60.8	100	90.6	59.4
	D.E.	0	14.3	0	0.29614	0.49899
Hypertree	Media	100	44.9	100	87.5	90.6
	D.E.	0	5.7	0	0.33601	0.29614
Relation	Media	100	59.8	100	90.6	43.8
	D.E.	0	13.4	0	0.29614	0.50402

- *Atención*: De acuerdo a los resultados que se presentan en la Tabla 11, la capacidad de percepción asociada con la *atención* presentó un alto grado de acierto para identificar la temática representada por cada una de los escenarios planteados. Esto significa que en todos los escenarios los participantes fueron capaces de identificar la temática planteada. Estos resultados permitieron identificar la importancia de la percepción visual para comprender la temática desplegada por cada nivel de jerarquía, así como la relación con la colección de recursos digitales; actividad que se llevó a cabo en la segunda fase del estudio (test de interacción- Sección 5.4).
- *Retención*: En cuanto a la capacidad de *retención*, a cada participante se le pidió que recordara por cada escenario el mayor número de términos posibles en un lapso de tiempo de 10 segundos. A mayor número de términos recordados, mayor sería la capacidad de retención. Las interfaces que favorecieron este proceso de retención de

términos en los participantes fueron *radial-search* ($M = 63,3$, D.E. = 12,9), *relation* ($M = 59,8$, D.E. = 13,4), e *icicle* ($M = 60,8$, D.E. = 14,3). Sin embargo, dentro de toda la muestra el análisis estadístico presentó una notable dispersión de la Desviación Estándar (D.E.), lo cual indica que fueron resultados no significativos para toda la muestra. Para verificar estos resultados, se llevó a cabo un proceso de revisión de videos a través de un experto en usabilidad para identificar posibles causas que reflejaron este resultado en los participantes, y uno de los factores que pudo haber influido en estos resultados se encontraba asociado a la familiaridad de las temáticas que encontraron algunos participantes, factor que podría haber favorecido la retención de un mayor número de términos.

- *Comprensión*: Los resultados asociados con la capacidad de *comprensión* mostraron una variación respecto a la *retención*, dado que los usuarios tenían que identificar la relación de asociación de los términos por cada nivel de profundidad representado en cada escenario. Por ejemplo, en la Figura 44 se presenta una taxonomía del escenario definido para la temática “*deportes*”. A cada usuario se le plantearon una serie de preguntas con el propósito de que lograra identificar i) la temática, y ii) comprender la relación asociativa que presentaban los términos por cada nivel de categoría, siendo: (nivel 1) deportes, (nivel 2) deportes de pelota y (nivel 3) deportes de equipo.



Figura 43. Ejemplo de niveles de relación para interfaz radial-search

Finalmente la Figura 44 muestra los resultados de los tres niveles de comprensión obtenidos por cada escenario, siendo el nivel 3 el de mayor grado de complejidad.

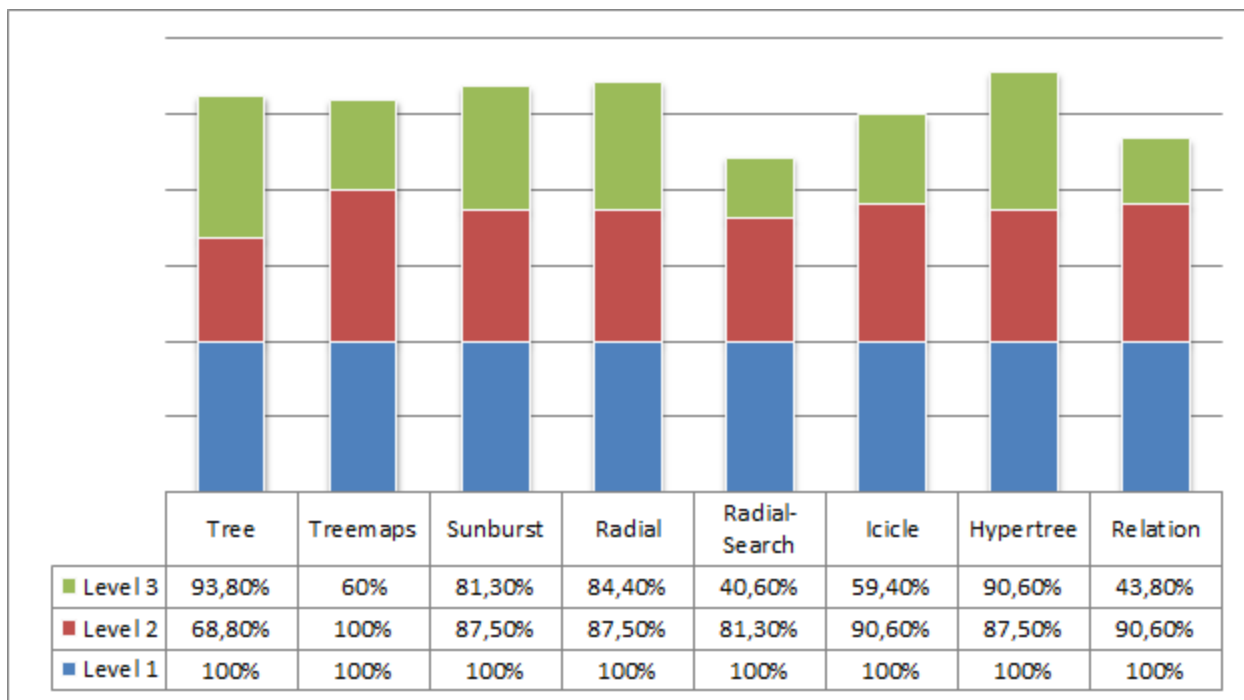


Figura 44. Nivel de comprensión por interfaces

Como resultado de este proceso, y de acuerdo a los resultados de la Figura 44, se encontró que el Nivel 1 presentaba un alto grado de comprensión (100%). De manera detallada los resultados de la Tabla 11, muestra que en el nivel 2 las interfaces que reflejaron un mayor grado de comprensión fueron *treemaps* (100%), *icicle* ($M = 90,6\%$, D.E. = 0,296) y *relation* ($M = 90,6\%$, D.E. = 0,296). Igualmente en la Tabla 11 muestra que para el nivel 3, las interfaces que presentaron un mayor nivel de comprensión fueron *tree* ($M = 93,80\%$, D.E. = 0,295), *hypertree* ($M = 90,6\%$, D.E. = 0,296) y *radial* ($M = 84,40\%$, D.E. = 0,368). Este criterio de análisis fue relevante, dado que permitió identificar por cada estructura de navegación las relaciones asociadas a cada nivel y categoría. Sin embargo, a pesar de la diferencia entre los niveles 2 y 3, estos resultados no fueron significativos, y tampoco fueron relevantes para los propósitos del estudio, dada la preferencia que tuvieron algunos participantes sobre algunas temáticas representadas en varios escenarios posiblemente por la familiaridad de los ejemplos.

En cuanto a los aspectos relacionados a la taxonomía, los resultados de esta prueba mostraron que los participantes prefieren interfaces cuya representación gráfica se despliegue a partir de estructuras definidas por categorías, reflejados en los resultados obtenidos por las interfaces *treemaps* e *icicle* (Figura 43). Sin embargo, al realizar una sesión de “*thinking-aloud*”, los participantes revelaron que estas preferencias se debieron a: i) estructuras de fácil entendimiento a partir de los niveles de clasificación definidos (tres niveles), ii) el número de términos por cada nivel de clasificación favoreció identificar las relaciones de asociación, es decir, 4 términos para el nivel 2 y 8 términos definidos en el nivel 3, y iii) la familiaridad de la temática. Siendo este último aspecto el que más favoreció a los participantes recordar una mayor cantidad de términos, e identificar las relaciones de asociación definidos entre los tres niveles.

A pesar de que todos los participantes lograron identificar las temáticas definidas en todos los escenarios planteados, no todos los participantes fueron capaces de identificar la relación de asociación de términos que se presentaron entre el segundo y tercer nivel de profundidad. Sin embargo, estos resultados (capacidad de recuerdo) no eran relevantes para nuestro estudio, dado que el interés principal se centraba en que cada participante lograra conocer cada interfaz para que identificara su forma y representación de un conjunto de términos dentro de una estructura taxonómica, mas no la capacidad de recuerdo de un conjunto de términos asociados a una temática en particular. Por lo tanto, esta fase de reconocimiento de las interfaces les permitió desarrollar habilidades para llevar a cabo la segunda parte del estudio a nivel de interacción.

6.1.2 Resultados de interacción

Para llevar a cabo este estudio se realizó una distribución de participantes de acuerdo al número de interfaces con el fin de cubrir todos los aspectos de la evaluación y determinar la fiabilidad de los resultados (Nielsen, 1994). La Figura 45 presenta un ejemplo de una de las actividades solicitadas al participante por parte del moderador, para ubicar el término "Art deco" y los niveles de profundidad (4) que el participante debía recorrer para localizar el término. El nivel 1 es el nivel de partida asociado a la categoría que permite llegar al término solicitado. Se le pidió a los usuarios que partieran su procesos de exploración a partir de la temática "European region", con el fin de que lograran navegar a través de los nodos "European" (1), "European styles and periods" (2), "Modern European styles and periods" (3), y finalmente ubicar el término "Art deco" (4).

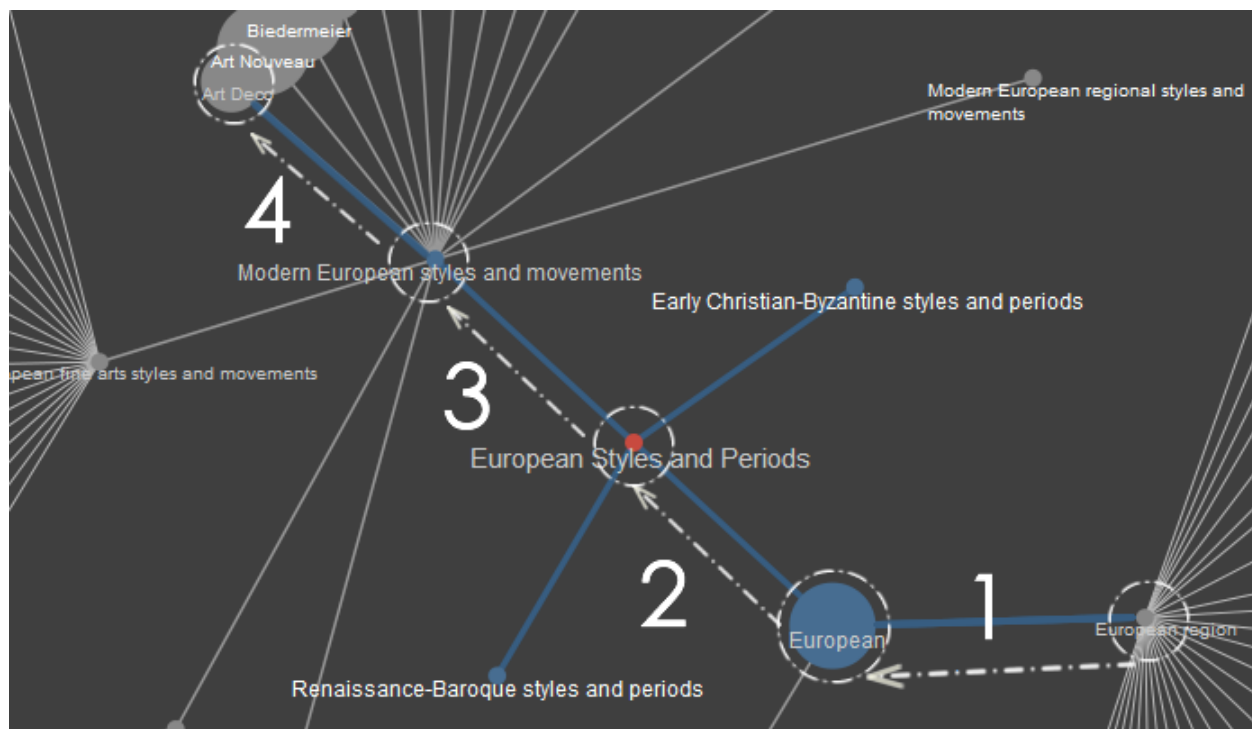


Figura 45. Niveles de profundidad para localizar el término "art deco" con la interfaz radial-search

Al final de esta prueba, los usuarios realizaron una valoración de varios aspectos relacionados con criterios de *interacción* y *usabilidad* (definidos en Sección 5.4). La Tabla 12 resume los resultados de esta prueba.

Tabla 12. Estadísticas sobre las impresiones del usuario relacionadas sobre aspectos de usabilidad

(1 = bajo nivel de satisfacción, 5 = alto nivel de satisfacción)

Interfaces	Estadística Descriptiva	Cobertura Temática	Clasificación	Navegación	Estética	Facilidad de Uso
Treemaps	Media	2.75	1.5	1.75	2.5	1.5
	D.E.	2.062	1	0.9574	1	0.5774
Tree	Media	4	4	3.5	2.75	4
	D.E.	1.414	0	0.5774	0.5	0
Sunburst	Media	2.25	2	2.25	2.75	2.25
	D.E.	1.893	1.414	1.258	0.9574	1.258
Relation	Media	3.5	2.75	3	3.25	2.5
	D.E.	1.291	0.9574	1.414	1.5	1
Radial-Search	Media	4	3.75	3.5	3.25	3.75
	D.E.	1.155	0.5	1	0.9574	0.5
Radial	Media	3.75	2.75	2.5	3.5	2.75
	D.E.	0.5	0.5	0.5774	1	0.5
Icicle	Media	4.25	3.5	4.5	3.25	4.25
	D.E.	0.5	1	0.5774	1.5	0.9574
Hypertree	Media	1	1	1	1.5	1
	D.E.	0	0	0	0.5774	0

La Tabla 12 presenta los resultados de las impresiones subjetivas de los participantes. De acuerdo a los resultados obtenidos a partir de la cobertura temática, las mejores interfaces fueron *icicle* (M = 4,25, D.E. = 0,5), *tree* (M = 4.0, D.E. = 1,414) y *radial-search* (M = 4.0, D.E. = 1,155). Similares resultados se presentaron en los niveles de clasificación y navegación.

A partir de las sesiones que se llevaron a cabo de “*thinking-aloud*”, y por medio de la revisión de vídeos con algunos de los participantes, se encontró que la mayoría de ellos se sintieron incómodos con las interfaces *hypertree* (M = 1 D.E. = 0,00) y *treemaps* (M = 1.5 D.E. = 0,57). Éstas fueron criticadas debido a la incapacidad que presentaron para proporcionar visualmente una ruta sencilla de navegación. Esta fue una de las razones principales por las cuales recibieron una de las valoraciones más bajas asociadas al atributo de “*facilidad de uso*”; resultado que también se vio reflejado en los demás atributos de usabilidad.

En cuanto a la interfaz *tree* ($M = 2,75$, D.E. = 0,5), fue mal valorada estéticamente debido a que la mayoría de opiniones de los participantes reflejaron impresiones subjetivas al calificarlos como interfaces "muy comunes y simples". Sin embargo, en otros atributos de usabilidad (facilidad de uso, clasificación y cobertura temática), esta interfaz tuvo una mayor aceptación.

Los resultados asociados al criterio de clasificación presentaron una baja evaluación sobre las interfaces *relation* ($M = 3,0$, D.E. = 1,414) y *sunburst* ($M = 2,25$, D.E. = 1,258). Las sesiones de "thinking-aloud" revelaron que estas valoraciones fueron ocasionados por los problemas que los participantes encontraron en la navegación, debido a la incapacidad que tenían estas interfaces para desplegar términos sobre altos niveles de jerarquía ($> = 6$ niveles de profundidad dentro de la estructura taxonómica). Adicional a esto, la interfaz *relation* presentó una curva de aprendizaje mayor para la comprensión de su esquema de navegación y localización de términos. Sin embargo, los peores resultados se identificaron en las interfaces *treemaps* ($M = 1,75$, D.E. = 0,957) e *hypertree* ($M = 1,0$, D.E. = 0,0), debido a la complejidad en que estas interfaces desplegaban las categorías al usuario, por lo tanto el proceso para localizar un término resultaba ser una actividad compleja de realizar. Por ejemplo, cuando los participantes navegaron dentro de una categoría que vinculaba más de cinco términos a través de la interfaz *hypertree*, los nodos de la interfaz se solapaban debido al efecto que presentaba esta interfaz mediante la técnica "ojo de pez", exigiendo un mayor esfuerzo por parte de los participantes mayoría de los participantes que usaron esta interfaz. Por lo tanto, esta fue una de las razones principales que llevaron a cabo el planteamiento de un estudio complementario asociado a la *precisión* de las interfaces que facilitaron la localización de términos (Sección 5.5) para localizar los términos solicitados.

6.1.4 Resultados de precisión

Cuando se compararon los resultados de las evaluaciones subjetivas con los resultados obtenidos de las actividades que cada participante debía realizar en las pruebas (Sección 5.4), surgieron situaciones donde algunos participantes tuvieron dificultades para identificar el número de recursos digitales asociados a los términos. Los participantes, en algunos casos comentaron su frustración por no encontrar los términos que el moderador solicitaba. La Figura 46 presenta los resultados de las interfaces que obtuvieron un mayor porcentaje acierto (interfaces que facilitaron el proceso de localización de términos) y fracaso (interfaces que no facilitaron el proceso de localización de términos durante las pruebas).

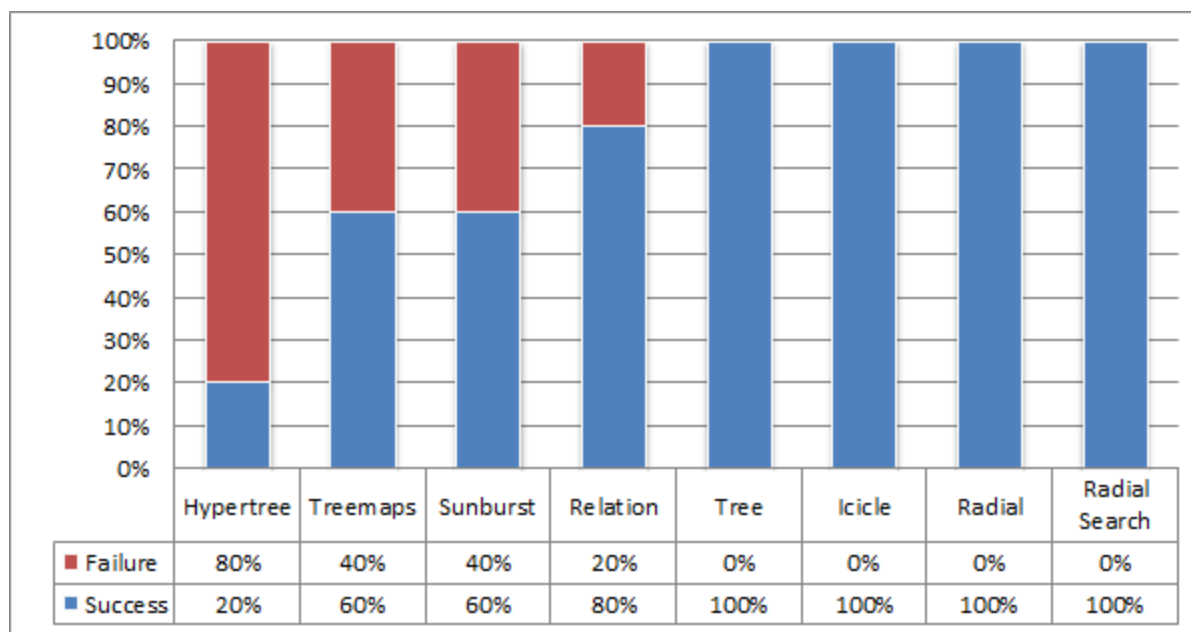


Figura 46. Porcentaje de éxito y fracaso para la localización de términos en interfaces

A pesar de llevar a cabo un proceso de aprendizaje en la etapa preliminar del estudio (fase de entrenamiento - Sección 5.1.3), se observó una tasa de fracaso (participantes que no lograron localizar el término solicitado) de todas las interfaces del 20%. Las interfaces con las tasas más altas de fracaso fueron *hypertree* (80%), *sunburst* (40%), *treemaps* (40%) y *relation* (20%). La interpretación realizada, de acuerdo con las sesiones realizadas de “*thinking-aloud*”, es que los participantes que interactuaron con la interfaz *hypertree* no tuvieron éxito en la localización de los términos debido a la distribución que realizaba de los nodos mediante el efecto “ojo de pez”, un efecto que impide una clara visión general de todos los términos dentro de un mismo nivel de taxonomía. Por otro lado, los participantes que utilizaron la interfaz *sunburst*, también tuvieron dificultades debido a que la interfaz desplegaba sobre la misma pantalla todos los términos junto con su estructura taxonómica. Esto produjo una dificultad al localizar términos debido a la saturación visual de los elementos en pantalla. Por último, en cuanto a la interfaz *treemaps*, los problemas que percibieron los participantes se relacionaron con los cambios bruscos de posición que presentaba la interfaz mediante su estructura de navegación al seleccionar un término, lo cual lograba desorientar totalmente al participante dentro de la estructura de navegación.

6.1.5 Resultados sobre percepción de la cobertura temática

Dada la relevancia de los resultados obtenidos en cuanto a percepción e interacción, se planteó un estudio complementario para verificar los resultados asociados a la cobertura temática, con el propósito de analizar si este criterio visualmente influía sobre el acceso a una colección de recursos digitales. De manera preliminar los resultados mostraron que las interfaces que presentaron visualmente diferentes tamaños en sus nodos, permitieron captar una mayor

atención del participante para lograr identificar términos con mayor o menor número de recursos. Estos efectos visuales facilitaban a los participantes identificar las áreas que presentaban una mayor o menor cobertura dentro de una estructura jerárquica. Criterio que les facilitó el proceso de acceso a una colección de recursos digitales. Para validar estos resultados preliminares, se llevó a cabo un análisis estadístico mediante análisis de varianza para identificar si existían diferencias significativas en el uso de interfaces con el fin de mejorar los procesos de acceso a una colección de recursos digitales a partir de su cobertura temática. En la Tabla 13, se presentan los resultados de este análisis.

Tabla 13. Promedio estadístico de valoración de interfaces a partir de cobertura temática

Interfaces	No. Interfaces	Media	Desviación Estándar (D.E.)
Hypertree	4	1.00	0.00
Icicle	4	4.25	0.5
R-Search	4	4.00	1.15
Radial	4	3.75	0.50
Relation	4	3.50	1.29
Sunburst	4	2.25	1.89
Tree	4	4.00	1.41
Treemaps	4	2.75	2.06
Total	32	3.18	1.55

Los resultados de la Tabla 13 presentan la media de los valores subjetivos de la cobertura temática de acuerdo a la percepción de 32 usuarios (4 usuarios por cada interfaz $n = 4$), y sus desviaciones estándar. Estos resultados indican una diferencia entre las medias de todas las interfaces, siendo representativos los resultados obtenidos por las interfaces *icicle* (4.25), *radial-search* (4.0) y *tree* (4.0). Para verificar si estos resultados de *cobertura temática* visualmente influyen sobre el acceso a una colección de recursos digitales, en la siguiente sección se lleva a cabo un proceso de verificación y contraste de hipótesis mediante un análisis de correlación entre todos los atributos y regresión lineal entre los resultados obtenidos de la *precisión* para el acceso a recursos digitales (resultados objetivos) a partir de los resultados de *cobertura temática* (resultados subjetivos) realizado por los participantes.

6.1.6. Comprobación de hipótesis

La hipótesis de estudio se orienta en determinar si existen diferencias significativas en el uso de interfaces gráficas, para facilitar el acceso a una colección de recursos digitales a partir de esquemas de representación de conocimiento definidos por su cobertura temática. A continuación se presenta su definición:

Hipótesis nula H_0 : La cobertura temática no facilita el acceso a recursos digitales

Hipótesis alternativa H_1 : La cobertura temática facilita el acceso a recursos digitales

Para identificar qué interfaces facilitaron la visualización de la cobertura temática se llevó a cabo una prueba de homogeneidad mediante un análisis de contraste, conocida como prueba de homogeneidad de grupos post-hoc utilizando la técnica de diferencias honestamente significativas de Tukey (HSD Tukey). Esta prueba se aplicó debido al gran número de comparaciones reflejados sobre las ocho interfaces. Los resultados de este análisis se presentan en la Tabla 14.

Tabla 14. Interfaces de prueba de homogeneidad utilizando HSD - técnica de Tukey

Interfaces	No Interfaces	Subconjunto para valor de alfa = 0.05	
		1	2
Hypertree	4	1.00	
Sunburst	4	2.25	
Treemaps	4	2.75	
Relation	4	3.5	
Radial	4		3.75
Tree	4		4.00
Radial-Search	4		4.00
Icicle	4		4.25
P-valor		0.075	0.35

El análisis de contraste realizado en la prueba de Tukey presentó dos grupos (datos proporcionados en la Tabla 14). En un primer grupo el estudio desplegó un conjunto de interfaces homogéneas que no facilitaron la identificación del número de recursos digitales a partir de la visualización de su *cobertura temática* (*hypertree*, *sunburst*, *relation* y *treemaps*). El resultado de este primer grupo indica ($p\text{-valor} = 0.075 < \alpha = 0.05$), un valor que define un nivel de significancia menor ($p\text{-valor} = 0.075$) al establecido en la prueba ($\alpha = 0.05$), lo cual estadísticamente indica que podemos aceptar la hipótesis. Sin embargo, en el segundo grupo se identificaron diferencias significativas para las interfaces (*radial*, *tree*, *radial-search* e *icicle*), lo que puede indicar que son interfaces aptas que facilitan el acceso a recursos digitales a partir del uso de esquemas de representación de conocimiento basados en la cobertura temática.

A partir de los resultados de la Tabla 14, se procedió a comprobar nuestra hipótesis mediante un análisis de correlación de Pearson, para determinar si la percepción visual de la *cobertura temática* tiene una relación sobre el acceso a una colección de recursos digitales a partir de aspectos como la *navegación*, *estética*, *facilidad de uso*, *clasificación* y *precisión*. Al definir el

análisis de correlación de Pearson, se verificó inicialmente si las muestras presentaban una distribución homogénea mediante la prueba de Kolmogorov-Smirnov, dando como resultado una distribución normal con un nivel de significancia ($p > 0.05$) para cada una de las muestras seleccionadas como se observa en la Tabla 15.

Tabla 15. Prueba distribución normal de Kolmogorov-Smirnov sobre todas las muestra

		Cobertura temática	Facilidad de Uso	Clasificación	Navegación	Estética	Precisión
Parámetros normales^{a,b}	Media	3.188	2.750	2.656	2.750	2.844	2.84
	Desviación típica	1.554	1.295	1.260	1.320	1.110	1.297
Diferencias más extremas	Absoluta	0.231	0.177	0.232	0.203	0.276	0.148
	Positiva	0.170	0.162	0.187	0.158	0.276	0.148
	Negativa	-0.231	-0.177	-0.232	-0.203	-0.195	-0.142
Z de Kolmogorov-Smirnov		1.305	0.999	1.312	1.150	1.563	0.84
Nivel de Significancia asintótica (bilateral)		0.066	0.272	0.064	0.142	0.051	0.481

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Al verificar la distribución normal de las muestras, se procede a realizar el análisis de correlación. A modo resumen, la Tabla 16 presenta los resultados del análisis de todos los atributos de usabilidad correlacionados a partir del criterio de evaluación *cobertura temática*.

Tabla 16. Análisis estadístico de atributos mediante correlación de Pearson (r)

Atributos de usabilidad	Cobertura temática	Facilidad de Uso	Clasificación	Navegación	Estética	Precisión
Cobertura temática	1					
Facilidad de Uso	0.729***	1				
Clasificación	0.726***	0.894***	1			
Navegación	0.700***	0.906***	0.723***	1		
Estética	0.354*	0.398*	0.329*	0.413*	1	
Precisión	0.639***	0.609***	0.578***	0.560***	0.475**	1
**. Correlación con nivel de significancia de 0.01 (bilateral)						
*. Correlación con nivel de significancia de 0.05 (bilateral)						

De acuerdo con la correlación de Pearson, el coeficiente (r) es un índice que mide la relación entre dos variables aleatorias. El valor máximo es 1 y a partir de 0.5 se considera un resultado aceptable y positivo para determinar que existe una relación entre las variables de estudio. Al analizar los resultados de la Tabla 16, se presentó una correlación moderada y positiva entre la *cobertura temática* y la *facilidad de uso* ($r = 0.729$), entre la *cobertura temática* y *clasificación* ($r = 0.726$) y entre la *cobertura temática* y la *navegación* ($r = 0.700$). Sin embargo se presentó una relación muy baja entre la *cobertura temática* y *estética* ($r = 0.354$). Cabe destacar que las correlaciones más fuertes se presentaron entre los atributos asociados con la *navegación - facilidad de uso* ($r = 0.906$) y la *clasificación - facilidad de uso* ($r = 0.894$).

De manera particular, los resultados del análisis de correlación entre la *cobertura temática* y *precisión* fue moderada, positiva y considerablemente significativa ($r = 0.639$). Sin embargo, para complementar estos resultados y cuantificar el grado de relación lineal existente entre estas dos variables (*cobertura temática* y *precisión*), a continuación en la Tabla 17 se llevó a cabo un análisis de varianza ANOVA, donde se evidenció nuevamente una diferencia significativa sobre el valor de p ($p\text{-valor} = 0.010 < \alpha = 0.05$).

Tabla 17. Coeficiente de regresión lineal para cobertura temática

Coeficiente ^a						
Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típico	Beta		
1	(Constante)	1.144	0.415		2.757	0.010
	Cobertura Temática	0.533	0.117	0.639	4.547	0.000
a. Variable dependiente: Precisión						

Para modelar la relación entre la *cobertura temática* y *precisión* la Tabla 17 presenta valores de los coeficientes de regresión lineal. Se presenta el coeficiente de la pendiente de la recta (Beta = 0.639), y el valor origen de la recta de regresión (B = 0.533). Por lo tanto, la regresión lineal que mejor representa este modelo lo podemos ver en la ecuación de regresión lineal presentada en la Ecuación 1.

Ecuación 1. Ecuación de regresión lineal de cobertura temática respecto a la precisión

$Y_1 = B_0 + B_1X_1$ $\text{Precisión} = 1.144 + 0.533 (\text{cobertura temática})$

La Ecuación 1, muestra el coeficiente (B_1) que representa la pendiente de la recta: cambio medio que se produce en los resultados de la *precisión* (Y_1), por cada valoración obtenida de la *cobertura temática* (X_1). El coeficiente (B_0) es el punto en el que la recta corta el eje vertical: el valor medio de la *precisión* que corresponde con el porcentaje de los resultados obtenidos de la

cobertura temática. Conociendo los valores de estos coeficientes, se puede reproducir la recta y describir la relación existente entre la valoración de la *cobertura temática* y los resultados de *precisión* al localizar el número de recursos digitales asociados a cada término. La Figura 47 presenta la representación gráfica del modelo de regresión lineal calculado.

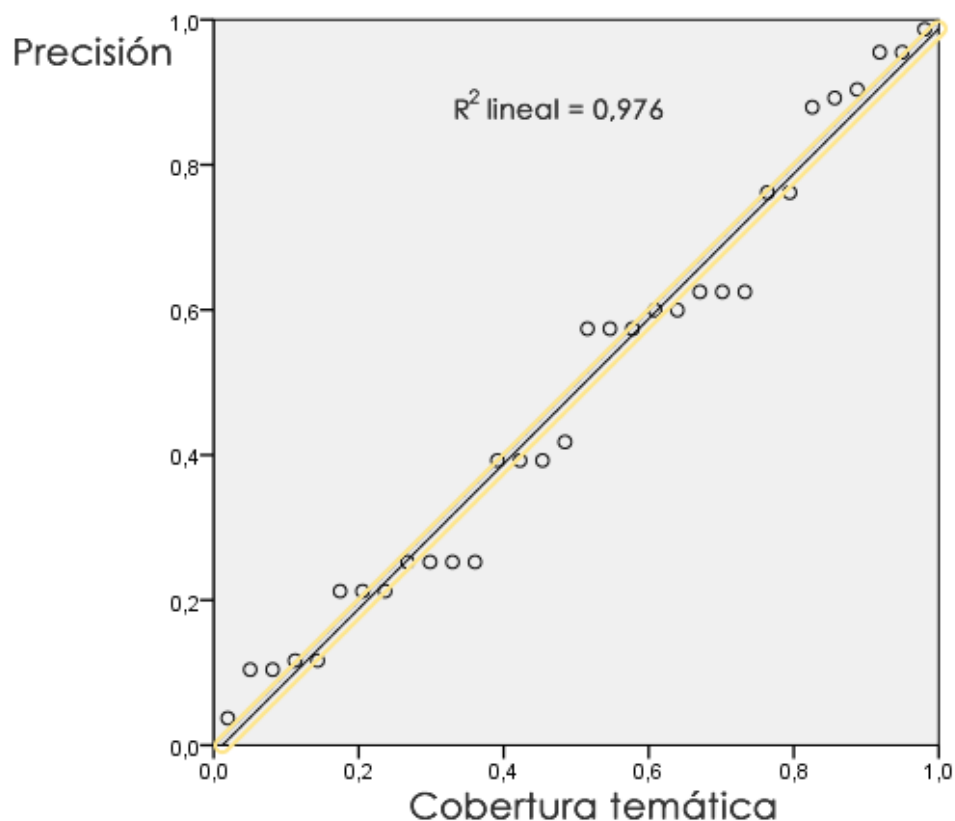


Figura 47. Regresión lineal entre cobertura temática y precisión

El coeficiente de determinación (R^2) que se muestra en la Figura 47, indica una medida estandarizada que toma valores entre 0 y 1 (0 cuando las variables son independientes y 1 cuando entre ellas existe relación perfecta). Este resultado indica que si una interfaz visualmente despliega una buena *cobertura temática*, esta condición facilitaría el acceso a una colección de recursos digitales con un 97.6% de fiabilidad de los resultados. De lo que se puede deducir que visualmente la *cobertura temática* (de acuerdo a las condiciones con que se llevaron a cabo las pruebas) presenta diferencias significativas para el acceso sobre una colección de recursos digitales, y por tanto es una condición que puede facilitar la ubicación de términos dentro de una estructura de navegación con un grado de relación de *cobertura temática* - *precisión* ($r = 0,639$ - ver Tabla 16) lo cual conduce a aceptar la hipótesis planteada.

Sin embargo, estos atributos de usabilidad no son condiciones suficientes y necesarias que garanticen el acceso y fiabilidad a una colección de recursos digitales. En el siguiente apartado se lleva a cabo un análisis detallado sobre otro aspecto definido como factor crítico dentro de los criterios de búsqueda y acceso a materiales relevantes, esto es la calidad de los metadatos de recursos digitales.

6.2 Análisis de calidad de los metadatos

Concluido el estudio de usabilidad, se procede a realizar un estudio paralelo de la calidad de metadatos, dada la variedad de resultados presentados en la Sección 4.2.3, mediante el proceso de análisis de cobertura de recursos digitales de Europeana asociados a la rama de conocimiento “*styles and periods*” definido en el tesauro AAT.

La calidad de los metadatos es una condición esencial para obtener resultados de búsquedas relevantes a partir del uso de motores de búsqueda definidos por ejemplo dentro de un repositorio (De la Prieta & Gil, 2010; Muñoz, et al., 2010); por lo tanto se considera como un elemento importante para mejorar las estrategias de indexación de objetos de aprendizaje almacenados en ellos (Ochoa, et al., 2005; Stuckenschmidt, et al., 2004; Wiley, 2002a). La omisión de metadatos constituye una mala clasificación de recursos digitales y por tanto es un factor clave que afecta los resultados de búsqueda a partir de criterios de búsqueda definidos por ejemplo dentro de un área de conocimiento específico.

Llevar a cabo un proceso de evaluación para medir la calidad de los contenidos es un indicador que permite evaluar los recursos digitales, sobre este escenario de trabajo, existen varios estudios que se refieren al diseño de métodos para evaluar la calidad de los recursos digitales (Chuanjun, 2004; Downes, 2007; Gonçalves, et al., 2007; Sanz, et al., 2009), así como la calidad de los contenidos existentes en las colecciones de recursos digitales (Chuanjun, 2004; Downes, 2007). A partir de estos métodos, se definen una serie de criterios de valoración para llevar a cabo evaluaciones de calidad de metadatos sobre una colección de recursos digitales. A continuación se presenta los resultados más representativos del análisis de calidad de metadatos que se llevó a cabo a partir de los recursos digitales extraídos en el primer proceso de exploración (556.000). Lo que se pretende con este análisis es analizar si la calidad de los metadatos favorece la clasificación de recursos digitales mediante criterios de búsqueda definidos por una temática o área de conocimiento. Es así, como este estudio pretende complementar el análisis de *completitud* (realizado en la Sección 4.3), y realizar un estudio de *precisión* de los metadatos definidos sobre recursos digitales explorados. De acuerdo a (Strong, Lee, & Wang, 1997), la calidad de los datos no se puede evaluar independientemente de las personas que utilizan los datos (los consumidores de datos). Para analizar la calidad de metadatos, se han propuesto una serie de criterios para clasificar la calidad en base a la *completitud*, *precisión*, *consistencia*, *coherencia*, *defecto de registros* (Barton, Currier, & Hey, 2003; Bruce & Hillmann, 2004; R. Wang & Strong, 1996). A continuación se presentan los resultados obtenidos del estudio de *completitud*, y seguidamente los resultados asociados al estudio de *precisión*.

6.2.1 Completitud

En el apartado 4.2.4, donde se definieron los criterios de búsqueda, se llevó a cabo un estudio preliminar sobre la *completitud* de los metadatos extraídos en el primer proceso de extracción de recursos digitales. En esta sección se lleva a cabo el estudio de *completitud* a partir de la

clasificación realizada en el modelo de intercambio de datos de Europeana EDM (Doerr, et al., 2010), con el fin de analizarlos a partir de un conjunto de metadatos *obligatorios*, *recomendados* y *opcionales*.

Tomando como base el concepto de *completitud* presentado en la Sección 4.2.4, se ha llevado a cabo un estudio donde se analizó la *completitud* partiendo del hecho de validar aquellos campos cuyo contenido sea un valor no nulo (Moen, et al., 1997; Ochoa & Duval, 2006; Park, 2005, 2009). La Tabla 18, presenta los resultados obtenidos del análisis de *completitud* para el conjunto de metadatos *obligatorios*.

Tabla 18. Resultados de completitud sobre elementos obligatorios

Elementos Obligatorios	Recursos Digitales	
	Registros Llenados	%
dc:title	547,98	98.47%
dc:Description	280,467	50.40%
dc:language	57,707	10.37%
europaana:dataProvider	0	0.00%
europaana:isShownAt	523,06	93.99%
europaana:isShownBy	263,924	47.42%
europaana:provider	556,514	100.00%
dc:subject	446,337	80.20%
dc:type	527,807	94.84%
dc:coverage	36,253	6.51%
dcterms:spatial	152,677	27.43%
europaana:rights	401,373	72.12%

Como su nombre lo indica, este conjunto de metadatos definidos como "*elementos obligatorios*", son metadatos que Europeana ha definido como indispensables para que proveedores de contenidos los diligencien en su totalidad. Partiendo de este propósito, los resultados de la Tabla 18 indican un bajo nivel de *completitud* con un porcentaje medio de 62.71% de manera global para todo este grupo de elementos de metadatos. Sin embargo, los elementos de metadatos definidos por proveedores externos como Dublin Core son los que en su totalidad mantienen un menor nivel de completitud con un promedio de 56.81% del total de metadatos. Promedio que altera el resultado para aquellos metadatos definidos por Europeana, que en su conjunto indican un porcentaje de 69%, resultado que sigue siendo todavía bajo, dada la relevancia de este conjunto de metadatos. A continuación en la Tabla 19 se describen los resultados para los metadatos definidos dentro de la categoría de elementos recomendados.

Tabla 19. Resultados de completitud sobre elementos recomendados

Elementos Recomendados	Recursos Digitales	
	Registros Llenados	%
dcterms:alternative	71,293	12.81%
dc:creator	348,142	62.56%
dc:contributor	114,476	20.57%
dc:date	276,118	49.62%
dcterms:created	77,638	13.95%
dcterms:issued	97,69	17.55%
dcterms:temporal	24,345	4.37%
dc:publisher	512,922	92.17%
dc:source	358,713	64.46%
dcterms:isPartOf	255,67	45.94%

Para el caso de los resultados obtenidos del conjunto de "*elementos recomendados*" presentados en la Tabla 19, los resultados del análisis muestran un promedio de 38.40% de *completitud* para todos los metadatos asociados a esta categoría, siendo *dc:publisher* el elemento con mayor grado de completitud (92.17%).

Tabla 20. Resultados de completitud sobre elementos opcionales

Elementos Opcionales	Recursos Digitales	
	Registros Llenados	%
dc:format	197,246	35.44%
dc:identifier	543,221	97.61%
dcterms:extent	168,786	30.33%
dcterms:medium	74,211	13.33%
dc:rights	338,706	60.86%
dcterms:provenance	9,267	1.67%
dc:relation	244,22	43.88%

Para el caso del conjunto de metadatos dentro de la categoría "*elementos opcionales*" que se presenta en Tabla 20, el nivel de completitud fue aún más bajo, con un promedio de 34.66% del total de los elementos representados para esta categoría. Por tanto, es claro que el nivel de *completitud* de metadatos disminuye a medida que la definición de metadatos no se define como una actividad obligatoria dentro del modelo de intercambio de datos definidos por EDM. Por tanto, el mayor número de elementos de metadatos clasificados dentro del conjunto de "*elementos obligatorios*" ofrecieron una mejor calidad de *completitud* respecto a los elementos

de metadatos clasificados en “*elementos recomendados*” y “*elementos opcionales*”. Aunque los resultados de manera global indican que para esta muestra extraída de Europeana la calidad de metadatos respecto a la *completitud* es muy baja.

Esta deficiencia en la *completitud* de metadatos, indican una limitación dentro de procesos de búsqueda de recursos digitales, dado que los metadatos facilitan el proceso de localización de un recurso digital a partir de las descripciones que se pueden tener de ellos. Sin embargo, no es suficiente realizar este tipo de estudios de *completitud* de manera aislada, dado que los campos de un metadato pueden contener cualquier tipo de valor. Por tal motivo, se ha complementado mediante un estudio asociado a la *precisión*, con el fin de determinar si la variada lista de resultados no relevantes de recursos digitales explorados en la Sección 4.2.3, fue influenciada por la mala definición de sus metadatos, y por consiguiente se obtienen recursos digitales no asociados al área de conocimiento definido.

6.2.2 Precisión

Para este análisis se orientó a partir de criterios de evaluación definidos por (Reddy & Wang, 1995) y Strong, et al (1997). De acuerdo a (R. Wang & Strong, 1996), la precisión es una dimensión definida para clasificar los recursos digitales de acuerdo a las capacidades intrínsecas de calidad de los datos, por tanto se relaciona con la coherencia de la descripción que se define en el metadato y en consecuencia lo que realmente representa el recurso digital.

En este caso de estudio, se han realizado varias comprobaciones de análisis de acuerdo a criterios de valoración definidos por Wang (1996). Por lo tanto, se ha realizado una adaptación orientado a verificar la descripción de metadatos en función de los recursos digitales desplegados. Por ejemplo, i) ¿el título se repite en otros recursos digitales que no están relacionados?, ii) ¿El título del recurso coincide con los recursos digitales asociados?, iii) ¿El título coincide con la descripción de los metadatos del recurso?, iv) ¿el idioma definido en los metadatos es el mismo que se utiliza para describir el recurso?, y, finalmente, v) ¿la descripción del tipo de formato del recurso, es coherente al que se despliega en el recurso?

La Tabla 21 presenta los resultados asociados a la *precisión* de calidad de metadatos para un conjunto de 5.500 recursos digitales asociados sobre elementos de metadatos más comunes usados en Europeana: *title*, *relation*, *description*, *type*, *language* y *subject*. Estos recursos fueron seleccionados al azar y de manera proporcional, a partir de los 556.000 recursos digitales que fueron extraídos en el primer proceso de exploración (Sección 4.2.3), con el propósito de abordar el mayor número de proveedores de contenidos (12 en total). Esto permitió tener una muestra representativa de cada proveedor de contenidos, a partir del total de recursos digitales explorados en Europeana.

Tabla 21. Resultados de precisión sobre recursos digitales seleccionados (5.500)

Elementos	% de Precisión	Número de recursos digitales precisos	% de Imprecisión	Número de recursos digitales imprecisos
Title	17.63%	986	82.37%	4608
Relation	7.24%	405	92.76%	5189
Description	3.41%	191	96.59%	5403
Type	6.20%	347	93.80%	5247
Language	5.65%	316	94.35%	5278
Subject	5.33%	298	94.67%	5296

El mayor problema que se encontró en este análisis es que los metadatos relacionados con el campo "*title*" se repetía para la mayoría de recursos digitales, dando una tasa de repetición del 82.37%. Por otro lado se presentaron inconsistencias asociadas con la ausencia en la definición del campo "*subject*", presentando tan solo un porcentaje de precisión de 5.33%. Se identificaron deficiencias en la definición de recursos digitales para el campo "*type*", con un porcentaje de precisión de 6.20%. Por último, el campo "*description*" fue el área de mayor deficiencia, con un porcentaje de precisión de apenas un 3.41%.

Los resultados de este estudio mostraron que algunos elementos de metadatos contenían los mismos valores (redundancia) con un porcentaje del 82% de registros repetidos para el elemento "*title*"; El 94% de los registros analizados indican una ausencia en la definición de los recursos digitales para el elemento "*language*" (ausencia de metadatos). El 93% de los registros contenían deficiencias en la definición de los recursos digitales que se contradecían respecto a las características del recurso digital (ambigüedad), y, finalmente, el elemento "*description*" fue el campo que presentó un mayor porcentaje de deficiencia, con un 96%. Este último resultado fue relevante para nuestro estudio, dado que la mayoría de las descripciones de los recursos digitales estaban ausentes en los recursos digitales, y además, porque las descripciones no correspondían con el idioma original definido el recurso digital (inconsistencia).

Todo lo anterior lógicamente influirá negativamente en los procesos de búsqueda. Por lo tanto la *precisión* es una condición necesaria que influye directamente sobre los resultados de búsqueda de recursos digitales. Dependiendo de la *precisión* con que se defina un recurso digital, se determinará el grado de relevancia de los resultados de búsqueda obtenidos, lo cual dentro de los criterios de búsqueda a partir de una temática o área de conocimiento, es un factor determinante para facilitar el acceso a cada uno de ellos. Por lo tanto estos resultados, junto con los obtenidos en los estudios de usabilidad, se detallarán y discutirán en mayor profundidad en la sección de discusiones.

6.3 Discusiones

En esta sección se presentan las discusiones relacionadas a los dos estudios realizados (usabilidad y análisis de metadatos). En primera instancia se presentan las discusiones relacionadas con el tema central de nuestro tema de investigación, asociadas con el diseño de interfaces a partir de técnicas de visualización como estrategia para facilitar el acceso a una colección de recursos digitales. A continuación se presentan las discusiones que dieron lugar al estudio complementario de análisis de calidad de metadatos de los recursos digitales explorados sobre la biblioteca digital Europea.

6.3.1 Discusiones de resultados de usabilidad

A partir de los resultados obtenidos sobre el estudio realizado asociado a las capacidades de percepción (Sección 5.2.1), se logró identificar que el nivel de retención no es una condición que afecta directamente sobre el reconocimiento de un área temática dentro de una estructura taxonómica (Tabla 11). Del mismo modo, la retención no afecta directamente sobre el razonamiento que necesita un usuario para identificar las relaciones de asociación en diferentes niveles de una estructura taxonómica, es decir, las relaciones temáticas que se vinculan sobre diferentes niveles en una taxonomía. Sin embargo, es importante mencionar que la desviación estándar de la retención (Tabla 11), mostró un alto nivel de dispersión. Por lo tanto, los resultados no fueron homogéneos. No obstante, este tipo de pruebas fue determinante para analizar el nivel de relevancia que presenta una temática, y las relaciones que representa dentro de una taxonomía. Por consiguiente, el análisis realizado sobre la satisfacción de cada participante de la impresión subjetiva (percepción visual) realizada a partir del uso de cada una de las interfaces en las pruebas de usabilidad, se evidenciaron como un factor importante asociado al papel que desempeña la visualización de una cobertura temática dentro de una jerarquía taxonómica. No obstante, los resultados del análisis de correlación sobre la cobertura temática (Tabla 16) evidenciaron una relación moderada y positiva sobre las pruebas objetivas, a partir de la precisión que tenían los participantes al ubicar los términos dentro de la interfaz visual de navegación, aunque no una relación fuerte entre el atributo asociado con la estética. Lo cual indica que la cobertura temática no es una condición determinante, pero si influye dentro del proceso de acceso a recursos digitales.

El proceso de análisis asociado a la prueba de percepción realizado por los participantes antes de la utilización de interfaces, facilitó la interacción de las interfaces asignadas para llevar a cabo procesos de búsqueda y exploración. Sin embargo, hay casos en que las evaluaciones realizadas en la prueba de percepción visual (primera parte del estudio, Sección 5.2.1) difieren respecto a las evaluaciones finales de prueba de interacción (segunda parte del estudio, Sección 5.2.2). Este fue el caso de interfaces *icicle* y *radial-search*, interfaces menos valoradas en cuanto a percepción (ver Figura 41). Sin embargo luego durante todo el estudio, estas interfaces fueron las mejores en los procesos de exploración y localización de términos de acuerdo a una cobertura temática al desplegar el número de recursos digitales (ver Tabla 11). Por otro lado es importante señalar que el uso de una buena estructura taxonómica, es una

condición que puede afectar el uso de interfaces visuales (Martín, et al., 2013). Por lo tanto, este factor fue uno de los motivos por el cual se llevó a cabo el uso de taxonomías sencillas al hacer las pruebas.

Las técnicas de visualización bajo estudio fueron en general percibidas por los participantes como útiles para mejorar el acceso a una colección de recursos digitales a través de una jerarquía taxonómica. Sin embargo, de acuerdo a opiniones realizadas por participantes en sesiones de “*thinking-aloud*”, se plantearon necesidades para complementar la navegación visual con la integración de búsquedas textuales. Los resultados de *facilidad de uso* (Tabla 12) y las pruebas de *precisión* (Figura 46) sobre la interfaz *radial-search* (la única interfaz con mecanismo de ayuda mediante búsqueda textual), confirman esta afirmación.

En resumen, se puede indicar que a los usuarios se les facilita un método que permita una rápida ubicación de un recurso en particular dentro de una estructura taxonómica, mediante el uso de criterios de búsqueda basados en la cobertura temática. Sin embargo, la curva de aprendizaje para comprender el uso de las interfaces visuales es uno de los factores, por no decir el único, que desempeña un papel importante para llevar a cabo un uso adecuado de este tipo de estrategias. Para lograr mantener esta curva de aprendizaje, es importante definir de manera estética, alternativas visuales que faciliten el entendimiento sobre la clasificación, localización y acceso a partir de un esquema de representación de conocimiento. Elementos que favorecerían la usabilidad de las interfaces, pero que se tendrían que evaluar dentro de tiempos de respuestas ideales para llevar a cabo consultas sobre un gran volumen de recursos digitales. Aspectos que no se evaluaron dentro de este estudio.

Por otro lado se puede mencionar que al analizar los resultados de usabilidad de manera individual, se identificó que las interfaces *hypertree*, *sunburst*, *treemaps*, *radial* y *relation* (ver Tabla 12) presentaron deficiencias para mostrar con eficacia términos de grandes colecciones de recursos digitales que se encuentren asociados dentro de altos niveles de profundidad en una estructura jerárquica. Por ejemplo, en la interfaz de *treemaps* la complejidad visual asociada al despliegue de términos de manera jerárquica se incrementaba a partir de la agrupación de términos definidos por los diagramas en bloques. Efecto que los participantes vieron como un obstáculo para ubicar fácilmente los términos solicitados en las pruebas. Sin embargo, las interfaces más efectivas (en relación a la *cobertura temática*, la *navegación* y *clasificación* - ver Tabla 12) fueron *icicle*, *tree* y la interfaz *radial* con buenos niveles de precisión.

Finalmente, como se abordó en la Sección 2.6, el tamaño de los datos sigue siendo un tema clave cuando se trata de visualizar información (Herman, et al., 2000; Huang, et al., 2008), y a partir de la experiencia al realizar diversas pruebas, se ha demostrado que una muestra bastante representativa de datos puede causar graves problemas, comprometiendo el rendimiento de la plataforma. Por ejemplo, el uso de las interfaces *sunburst* y *relation* los usuarios experimentaron dificultades (tiempos de carga inaceptables) al esperar un tiempo adicional para desplegar un proceso de consulta dentro de una jerarquía taxonómica. Por lo tanto, el tamaño y la profundidad de la estructura jerárquica se identificaron como un problema

para el despliegue de datos en pantalla. Aunque existen resultados de estudios que afirman que una interfaz de búsqueda mediante la técnica de visualización *treemaps* permite trabajar con miles de nodos de manera eficaz (Blanch & Lecolinet, 2007; Muelder & Ma, 2008), las pruebas realizadas en este estudio mostraron resultados diferentes en el nivel de desempeño, dado que los elementos de los niveles inferiores no eran fácilmente percibidos por su tamaño tan pequeño, lo cual lograba perder el contexto global de toda la jerarquía (ver Tabla 12).

6.3.2 Discusiones de resultados sobre análisis de calidad de metadatos

Si los campos de descripción de recursos digitales sobre cualquier repositorio no están bien descritos, esta limitación en los metadatos tiene un impacto muy negativo en los procesos de búsqueda de recursos digitales. Esto generaría que los recursos de alta calidad, no podrían ser ubicados fácilmente por parte de usuarios finales si no se encuentran bien definidos. Por lo tanto, la omisión del uso de metadatos en el caso en particular de Europeana, no facilitaría procesos de creación de objetos de aprendizaje a partir de la ausencia o mala definición de las descripciones de recursos digitales. Sin embargo, a pesar de que se encuentran realizando grandes esfuerzos para mejorar la calidad de sus metadatos mediante el modelo de representación de datos EDM, una posible solución sería la inclusión de nuevos campos de metadatos para clasificar eficazmente un recurso digital a partir de criterios de búsqueda definidos por temáticas o áreas de conocimiento. Esta actividad se puede complementar mediante la definición de metadatos LOM como "*classification*" y "*relation*". Estos elementos de metadatos adicionales proporcionarían una clasificación por áreas o temas, que podrían mejorar los resultados de búsqueda de recursos digitales de acuerdo a un área de conocimiento específico, y así mejorar los procesos de búsqueda y simplificar el acceso sobre una colección de recursos digitales haciendo uso de interfaces de búsqueda visual.

Los resultados obtenidos sobre el análisis de calidad lograron evidenciar una escasa completitud en la definición de metadatos clasificados dentro del modelo de datos de Europeana EDM. Sin embargo, este tipo de resultados no es un descubrimiento nuevo. En otros estudios sobre repositorios digitales como ARIADNE (Ternier et al., 2009), la biblioteca Nacional de Ciencias digital (Fox, Gonçalves, & Kipp, 2002) y otras colecciones de recursos digitales (Bui & Park, 2006) se han encontrado deficiencias en la definición de metadatos. De hecho, la deficiencia en la definición de metadatos es uno de los principales factores que influyen directamente sobre la búsqueda de recursos digitales, y más concretamente, la falta de *precisión* en la definición de elementos de metadatos. Del mismo modo, la ausencia de elementos de metadatos que describen la clasificación a partir de una temática o área de conocimiento de un recurso digital, se vieron reflejados en los pobres resultados de búsqueda planteados en el estudio de cobertura temática. Factores que también influyen en los resultados de búsqueda, y que han sido factores mencionados por Cechinel (2009).

A pesar de las deficiencias detectadas en la calidad de metadatos asociados a criterios de *precisión* y *completitud*, Europeana presenta grandes oportunidades de ser una biblioteca digital con posibilidades de reutilización de recursos digitales para el desarrollo de objetos de

aprendizaje. No sólo por los resultados de cobertura y el gran volumen de recursos digitales que dispone esta biblioteca, sino por el modelo de intercambio de datos EDM. Estrategia, que se perfila como un modelo que a futuro cubrirá totalmente aspectos relacionados a búsquedas semánticas, y por tanto facilitaría una mejor calidad en los proceso de búsqueda mediante la vinculación de los mismos a través de Linked Data (Bizer, et al., 2009; Dietze et al., 2012; Haslhofer & Isaac, 2011). Sin embargo, con el fin de mejorar la calidad de los mismos, Europeana debe hacer grandes esfuerzos, no solamente para mejorar la calidad de los metadatos de los recursos digitales, sino además, definir estrategias mediante la implementación de metadatos alternativos que faciliten la búsqueda de recursos digitales a partir de una temática o área de conocimiento. Para ello es fundamental la integración de esquemas de representación de conocimiento como el uso de ontologías o de tesauros, para clasificar los recursos digitales asociados a una rama de conocimiento y su posterior vinculación.

Para ampliar mejor las anteriores ideas, a continuación se presenta la elaboración de una propuesta que dentro del área de visualización de información permitirá identificar retos y oportunidades en el área de repositorios digitales, para el desarrollo de interfaces visuales que faciliten el acceso a una colección de recursos digitales.

6.4. Elaboración de una propuesta

De acuerdo a los resultados obtenidos de los estudios planteados en esta investigación y algunas recomendaciones planteadas sobre estudios realizados a partir del análisis de usabilidad en repositorios académicos Organic.Edunet y VOA3R (Martin, et al., 2013). El siguiente apartado tiene como propósito presentar una serie de recomendaciones, sugerencias, oportunidades y retos que presenta el área de la visualización para el desarrollo e implementación de interfaces de búsqueda sobre repositorios digitales.

Con el propósito de recopilar y resaltar los aspectos más relevantes, se ha definido un modelo representado en la Figura 48; en esta figura (48) se plantean cinco etapas que se toman en consideración para realizar un proceso de verificación, previo al desarrollo e integración de este tipo de estrategias de búsquedas sobre repositorios.



Figura 48. Etapas y consideraciones para el desarrollo de interfaces visuales a partir de esquemas de representación de conocimiento

6.4.1 Análisis y verificación de esquema de representación de conocimiento

Uno de los desafíos que se ven enfrentados la mayoría de repositorios digitales, se refleja no solo en la gestión de grandes volúmenes de recursos digitales, sino en el manejo adecuado de un vocabulario enriquecido que facilite la descripción de cada uno de ellos. Otra de las limitaciones pasa por la definición de formatos y lenguajes que permitan llevar a cabo una integración, a partir de esquemas de representación de conocimiento. Uno de los factores que influyen directamente sobre el desarrollo de este tipo de estrategias es el esquema de representación de conocimiento definido en el repositorio, normalmente ontologías y tesauros. Por lo tanto, es importante llevar a cabo un análisis del esquema de representación de conocimiento para identificar formalmente cuál es su propósito de uso, y qué papel desempeña dentro del repositorio a saber: i) favorecer procesos de búsqueda de recursos digitales, ii) uso de vocabulario enriquecido para describir formalmente cada una de las descripciones del recurso digital, iii) clasificación y niveles de jerarquías definidos en el esquema de representación de conocimiento y iv) relaciones directas e indirectas de vocabularios generados a partir de un conjunto especializado de lenguajes semánticos en formatos SKOS para la búsqueda de términos dentro de un esquema de navegación.

Un esquema de representación de conocimiento puede tener muchas implicaciones dentro de un proceso de clasificación y gestión de recursos digitales, y de la misma forma puede

influenciar de manera significativa sobre procesos de búsqueda que se definan a partir del uso de técnicas de visualización dentro de uno o varios repositorios digitales. Para lograr disminuir esta brecha de complejidad, es importante identificar una serie de pautas que el esquema de representación de conocimiento debe cumplir a la hora de llevar a cabo este tipo de implementaciones. La Figura 49 presenta la valoración de una serie de factores que se deben identificar de acuerdo a las facilidades de uso que pueden ofrecer tres esquemas de representación de conocimiento, como instrumentos que faciliten la implementación de estrategias de búsquedas visuales sobre repositorios digitales.

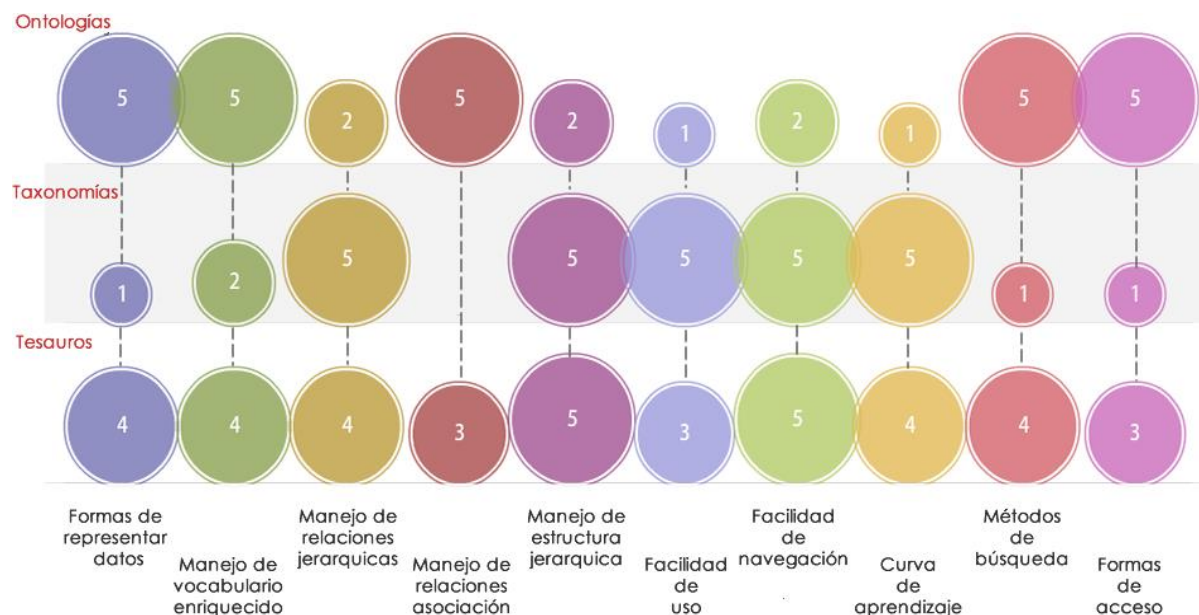


Figura 49. Criterios de evaluación para selección de esquema de representación de conocimiento

La Figura 49 presenta en el eje horizontal los factores que se han definido para llevar a cabo el proceso de comprobación de tres instrumentos de representación de conocimiento. Los primeros cinco factores se encuentran asociados a criterios de gestión y clasificación de información seleccionados a partir de estudios realizados por Zeng (2008) y Sánchez (2009). Los siguientes tres factores, se relacionan con criterios de usabilidad, y los dos últimos factores hacen énfasis a criterios de búsquedas y facilidades para su implementación dentro de un repositorio digital.

En la representación de la Figura 49 el primer grupo de factores (criterios de gestión y clasificación) indica que las ontologías presentan una ventaja significativa sobre el uso de *tesauros* y *taxonomías*. Aunque el grupo de factores asociados con criterios de usabilidad los *tesauros* presentan una valoración positiva, respecto al uso de *taxonomías* y *ontologías*. Sin embargo como se abordó en la Sección 2.1, las *taxonomías* se encuentran mejor representadas a partir de la implementación de *tesauros* que sobre *ontologías*. Aunque dependiendo del propósito del modelo de representación de conocimiento definido a través de una ontología, estas últimas favorecerían procesos de búsqueda basados en el enriquecimiento

semántico que proporcionen sobre una colección de recursos digitales, siempre y cuando se establezcan una serie de criterios de clasificación orientados a la ubicación de recursos digitales a partir de una temática o área de conocimiento específico. Proceso que puede tomar un tiempo extra de análisis e implementación sobre la definición del modelo ontológico, previo a la integración de interfaces de búsqueda navegacionales a partir del uso de técnicas de visualización.

6.4.2 Selección de técnicas de visualización

Para realizar un proceso de selección de técnicas de visualización, se ha planteado un esquema de tres categorías asociadas con: 1) características propias de cada técnica de visualización, 2) características de desarrollo, y 3) características de uso por parte de usuarios. La Figura 50 presenta una representación y características asociadas a cada factor de evaluación.

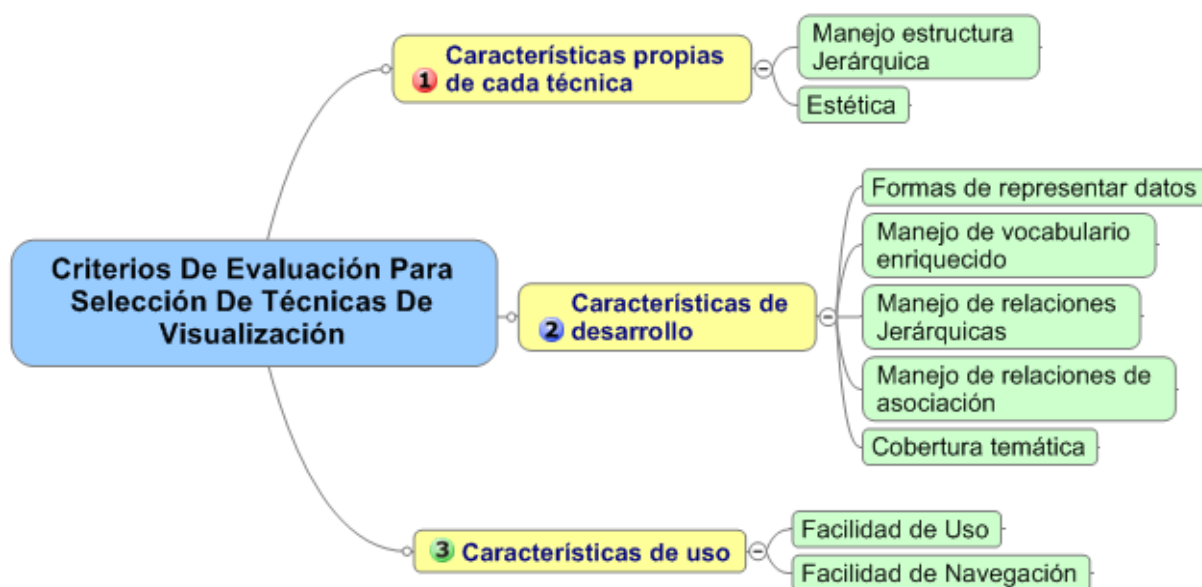


Figura 50. Criterios de evaluación para selección de técnicas de visualización

A partir de esta clasificación, la Figura 50 presenta a modo resumen una evaluación realizada para la selección de técnicas de visualización que faciliten el proceso de desarrollo de interfaces de búsqueda visual sobre repositorios digitales. El proceso de valoración realizado a cada interfaz se realizó de acuerdo a resultados de usabilidad obtenidos a lo largo de los estudios realizados, y la experiencia adquirida en la implementación de cada una de ellas en el desarrollo del proyecto.

Hypertree	2	2	2	2	2	2	1	3	3	1	1
Icicle	3	3	4	4	4	2	5	3	3	4	4
Radial	3	4	4	4	4	3	3	3	4	3	3
Radial-search	3	4	4	3	4	3	4	3	4	3	4
Relation	2	5	3	5	5	5	2	2	4	2	2
Sunburst	4	3	3	2	3	2	2	3	3	3	4
Tree	4	4	5	3	5	3	4	5	4	5	5
Treemaps	2	3	2	2	3	2	3	2	3	2	3
Manejo de estructura jerárquica	(1)	(2)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	
Estética											
Formas de representar datos											
Manejo de vocabulario enriquecido											
Manejo de relaciones jerárquicas											
Manejo de relaciones asociación											
Cobertura temática											
Curva de aprendizaje											
Soporte métodos de búsqueda											
Facilidad de uso											
Facilidad de navegación											

Figura 51. Resultados de evaluación para selección de técnicas de visualización

A partir de la valoración realizada en la Figura 51 se puede identificar las siguientes interfaces como las más adecuadas para implementar de acuerdo a los tres criterios de evaluación definidos:

- *Características propias de cada técnica de visualización:* Las interfaces que presentaron mejores valoraciones de acuerdo a propiedades estéticas y métodos de clasificación, a partir de las jerarquías definidas por el tesauro AAT fueron *relation*, *tree* y *radial*. Aunque la interfaz *relation* presentó una alta valoración asociado a su estética, la forma de representar las jerarquías por niveles desfavorecen notablemente su implementación.
- *Características de desarrollo:* Sobre esta categoría se encuentran las interfaces *tree*, *icicle* y *radial-search* como las más apropiadas para el despliegue de estructuras jerárquicas, manejo de vocabulario enriquecido, relaciones de jerarquías y de asociación. De igual forma, para la integración de métodos de búsqueda específicamente orientadas a la cobertura temática a partir de un área de conocimiento.
- *Características de uso:* Para esta categoría se identifican las interfaces *tree*, *icicle* y *radial-search* como las más opcionadas para facilitar el proceso de navegación e implementación a partir de un esquema de representación de conocimiento previamente establecido.

6.4.3 Selección de herramientas y bibliotecas de funciones gráficas

El desarrollo de aplicaciones basadas en visualización de información, requiere de componentes que faciliten su uso, adaptación y soporte a lo largo del periodo de vida de un proyecto. Como se indicó en la Sección 4.3.2, existe un gran listado de proyectos en Internet que facilitan la implementación de técnicas de visualización a partir de una serie de bibliotecas de funciones y APIs de tipo Open Source. Algunas se componen de funciones computacionales mediante el uso de estrategias de diagramación de tipo Vectores Gráficos Escalables SVG (basados en formas). Este tipo de diagramación ofrece grandes capacidades para el soporte de aplicaciones gráficas sobre la mayoría de navegadores en Internet y el desarrollo de aplicaciones interactivas sin la necesidad de usar tecnologías especializadas.

Otras tecnologías similares ofrecen capacidades de soporte para visualizar código sobre cualquier navegador compatible con funcionalidades HTML5 (W3C, 2013) y vinculación de etiquetas tipo *canvas* (basado en pixeles). Este tipo de tecnología mejoría, en cierta medida la experiencia del usuario, ya que omite la instalación de plugins o complementos, dado que la mayoría de navegadores proporcionan los gráficos en forma nativa facilitando la aceleración de gráficos por hardware (Boulos, Warren, Gong, & Yue, 2010; Fulton & Fulton, 2013). Por lo tanto para cualquier gráfico vectorial que se base en estas tecnologías, el programador y/o diseñador gráfico deberá centrar mayores esfuerzos dependiendo de la actividad a realizar.

El modo de operación de estas herramientas basadas en gráficos vectoriales, es trabajar internamente mediante la compilación de bibliotecas de funciones Open Source y convertir el código en scripting como por ejemplo mediante lenguaje JavaScript. Esto permite fácilmente a un navegador con soporte de HTML5 la generación dinámica de imágenes estáticas y animaciones por medio de scripting. Sin embargo, una de las diferencias relevantes entre SVG y HTML5, es que la primera necesita trabajar con mayor memoria para retener y manipular los resultados de los códigos para su procesamiento, mientras que HTML5 lo realiza de manera inmediata. (Fulton & Fulton, 2013; W3C, 2013).

A continuación en la Tabla 22, se presenta a modo resumen las características y soportes de algunas de las herramientas más representativas de tipo Open Source y orientados a la Web.

Tabla 22. Características de herramientas de visualización open Source

Proyecto	Lenguaje	Dependencias	Tecnologías	Soporte técnicas visualización					
				Tree	Treemaps	Icicle	Radial	Sunburst	Hyperbolic
Protovis	JavaScript	ninguna	HTML, SVG y CSS	si	si	si	si	si	si
D3js	JavaScript	ninguna	HTML5, canvas, SVG y CSS	si	si	si	si	si	si
Infovis	JavaScript	ninguna	HTML5, canvas y CSS	si	si	si	si	si	si
Processingjs	JavaScript	ninguna	HTML5, canvas y CSS	si	no	no	si	no	no

Proyecto	Lenguaje	Dependencias	Tecnologías	Soporte técnicas visualización					
				Tree	Treemaps	Icicle	Radial	Sunburst	Hyperbolic
Processing	Java	Java	HTML5, canvas, WebGL y CSS	no	no	no	no	no	no
Raphaël	JavaScript	ninguna	HTML, SVG y CSS	si	no	no	si	no	no
Prefuse	Java	Java	2D graphics library	si	si	no	si	si	no
Sigmajs	JavaScript	ninguna	HTML5, canvas y CSS	si	no	no	no	no	no
Arborjs	JavaScript	ninguna	HTML5, SVG, canvas y CSS	si	no	no	si	no	no
Moowheel	JavaScript	ninguna	HTML5, SVG, canvas y CSS	no	no	no	si	no	no
Chap	JavaScript	ninguna	HTML, SVG, canvas y CSS	si	no	no	si	no	no
Axxis	Flex	Flash	Flex, ActionScript	no	si	no	no	si	no
Kap Lab	Flex	Flash	Flex, ActionScript	no	si	no	no	si	no
Cytoscape	JavaScript	ninguna	HTML, SVG, canvas y CSS	si	no	no	si	no	no
Nvd3	JavaScript	ninguna	HTML, SVG, canvas y CSS	si	no	no	no	no	no
Miso	JavaScript	ninguna	HTML, SVG, canvas y CSS	si	si	no	no	no	no

A partir de la clasificación realizada en la Tabla 22, es importante mencionar tecnologías y arquitecturas de trabajo como Neo4j (Holzschuher & Peinl, 2013) basadas en NoSQL (Strauch, Sites, & Kriha, 2011); tecnología que define un entorno de trabajo basado en entornos de programación Javascript del lado del servidor como lo es Node (Tilkov & Vinoski, 2010) y el uso de librerías gráficas como D3js (Bostock & Heer, 2009). Este tipo de soluciones no se contemplaron en la presente investigación para su análisis e integración dada que son herramientas comerciales, y que adicionalmente presenta limitaciones para el número de nodos a trabajar en su versión de uso libre.

La figura 52 presenta una valoración realizada de cada una de las bibliotecas de funciones anteriormente mencionadas sobre aspectos relacionados a su desempeño, con el propósito de determinar criterios de selección definidas por (Bostock & Heer, 2009; Fekete, 2004; Jeffrey Heer, Card, & Landay, 2005; Smoot, Ono, Ruscheinski, Wang, & Ideker, 2011; Wylie & Baumes, 2009), y complementadas de acuerdo a la experiencia de uso en este proyecto, a partir de las necesidades requeridas para su implementación.

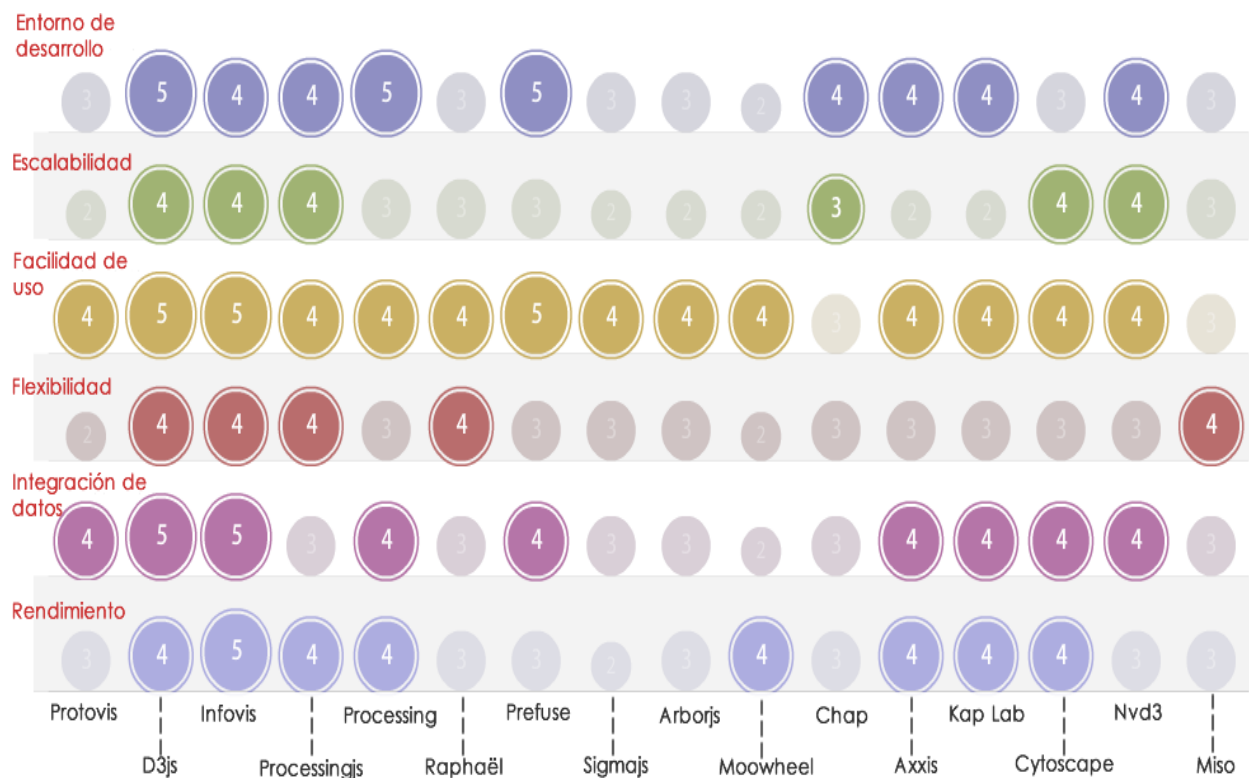


Figura 52. Criterios de valoración herramientas de visualización

6.4.4 Definición de framework de integración de búsquedas

Dentro de esta fase, solamente se mencionará la relevancia de la definición de un framework que facilite la integración de las estrategias de búsquedas navegacionales, con estrategias de búsquedas convencionales. Esta definición facilitaría a cualquier usuario, comprender la estructura de navegación empleada por la técnica de visualización, mediante la localización del término a consultar dentro de su estructura taxonómica. Lo cual favorecería las actividades de exploración y la curva de aprendizaje será menos rigurosa para los usuarios. Por lo tanto la definición de este framework debe contener por lo menos el soporte básico de módulos y componentes que diferencien la lógica del negocio y la del cliente, reflejados en arquitecturas de tipo Modelo Vista Controlador MVC o arquitecturas de tres Capas como se mencionó en la Sección 4.3.3.

La Figura 53 se propone una arquitectura en tres capas, donde se incluyen los componentes gráficos que se deben incluir a partir de arquitecturas de desarrollo robustas que se incluyan dentro del proyecto.

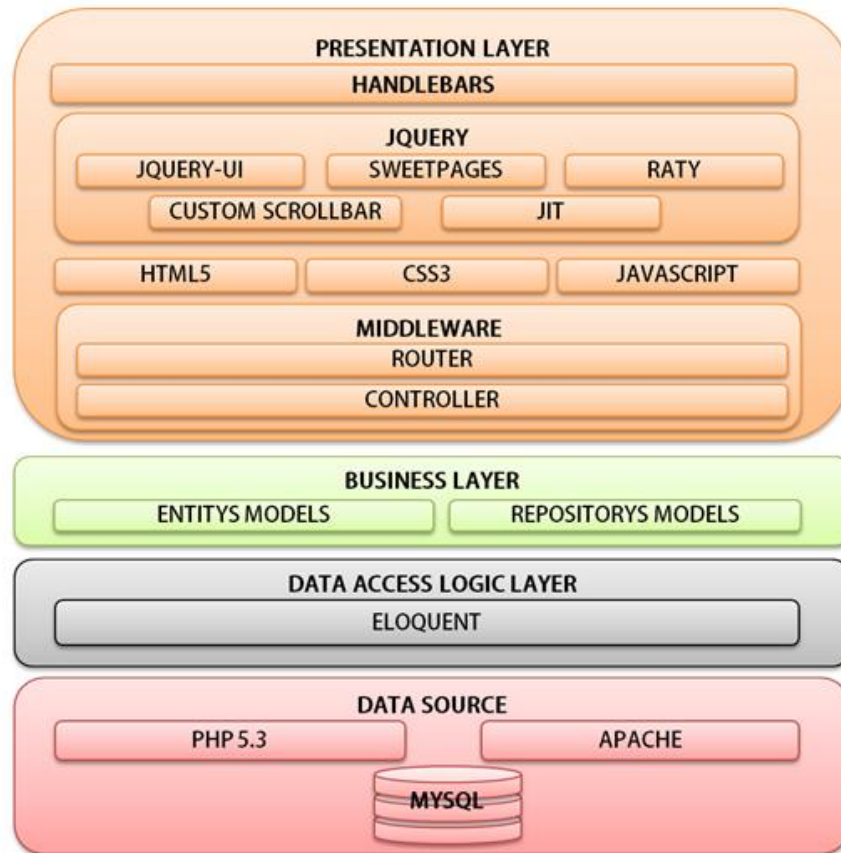


Figura 53. Arquitectura del modelo propuesto de tres capas para el área de visualización

- *Capa de almacenamiento y acceso a los datos:* Presenta tanto el almacenamiento como el tratamiento de los datos para el acceso a las colecciones de recursos digitales, la estructura del esquema de representación de conocimiento o vocabularios controlados que se definan para realizar procesos de búsqueda.
- *Capa de negocio:* Presentan las bibliotecas de funciones y herramientas de soporte para la integración de las búsquedas. Esta capa realiza peticiones para capturar la información que se desea consultar y gestionar. Esta capa se conforma por dos diferentes módulos que tienen como principal objetivo el tratamiento de los objetos. dentro de la aplicación gráfico para la implementación de técnicas de visualización. De la misma forma podemos integrar componentes complementarios que facilitan las actividades de búsqueda mediante el uso de bibliotecas de funciones gráficas.
- *Capa de presentación:* Presenta las estrategias que facilitan el uso de las bibliotecas de funciones y demás componentes visuales definidos en la capa de aplicación, a partir del uso de componentes visuales y de soporte definidos a través de HTML5. De la misma forma podemos integrar componentes complementarios que facilitan las actividades de búsqueda mediante el uso de bibliotecas de funciones gráficas. Esta capa es

implementada para el manejo de los datos y su presentación sobre la vista del usuario. Por lo tanto esta encargada del tratamiento de toda la representación visual de los datos, tanto en el lado del usuario como en los controladores.

6.4.5 Definición de despliegue de resultados y complementos de consultas

Finalmente este apartado pretende resaltar la importancia que cumplen los mecanismos que favorecen la carga de miles de recursos digitales. Este aspecto se ha incluido dentro de este listado de recomendaciones dado que es un factor importante que influye considerablemente en la etapa de selección del recurso digital. Hasta este punto se ha enfocado en los criterios que deben cumplir los esquemas de representación de conocimiento, las capacidades que pueden ofrecer las diferentes técnicas de visualización para implementar dentro de estructuras de navegación taxonómica, y se han definido algunas especificaciones para resaltar la importancia de la selección de un framework con soporte de bibliotecas de funciones gráficas para la vinculación de métodos de interacción visual sobre las estructuras de navegación. Sin embargo, se ha de incluir este último aspecto asociado con el despliegue de resultados, dada la gran necesidad de ofrecer alternativas de acceso efectivas dentro de un proceso normal de localización de recursos digitales relevantes por parte de un usuario. Por lo tanto, el despliegue de recursos digitales, es un aspecto importante que todo creador, administrador de repositorios digitales debe incluir, dado que es un momento dentro del proceso de búsqueda de recursos digitales, que puede tener un alto margen de deserción por parte de los usuarios.

Para lograr mitigar este riesgo, la siguiente sección se presenta un listado de estrategias y alternativas que facilitarían la culminación de un proceso de búsqueda a partir del uso de técnicas de visualización.

6.5 Desafíos y retos del área de visualización de información

Es claro que la visualización de información no puede abarcar toda la solución dentro de un proceso de búsqueda, pero su inclusión facilitaría el acceso mediante la implementación de diferentes técnicas y estrategias de búsqueda. En este sentido, el despliegue de recursos digitales, forma parte de las soluciones que un creador de repositorios digitales debe considerar para facilitar elementos básicos de usabilidad, para ello es importante resaltar los siguientes principios desde el punto de vista de la interfaz:

- Debe mostrar de manera clara y en un lugar visible el número de recursos digitales desplegados en el proceso de cada consulta.
- Debe disponer de un mecanismo de paginación que facilite la navegación y defina el número de páginas reales que despliega cada consulta.

- Debe desplegar de manera clara el número de recursos digitales desplegados por cada página de consulta a partir del método de paginación implementado.
- Cada recurso digital desplegado por un listado de paginación, debe contener información relevante asociado con el *título*, *descripción*, *proveedor de contenido*, vinculación del recurso digital con otros recursos similares bajo la misma área de conocimiento y en el mejor de los casos la pre-visualización del recurso digital (imagen, video).
- Complementar el despliegue de resultados mediante la implementación de mecanismos de valoración de recursos digitales, de acuerdo al perfil que mejor se adapte a las características del repositorio digital, mediante evaluación por expertos, usuarios registrados, profesionales, estudiantes, etc.
- Debe clasificar el acceso de los recursos digitales de acuerdo a las características de uso del recurso digital (Educación media, avanzada, continuada, universitaria, profesional), perfiles de usuarios (maestros, estudiantes, profesionales), formatos del recurso digital (imagen, video, documento rtf, presentación, filminas, etc.), idioma del recurso digital (inglés, español, etc.).

Es indudable las capacidades que puede ofrecer la visualización de información dentro del campo de los repositorios digitales, pero desafortunadamente también existen una serie de retos y desafíos, que a la fecha el área de la visualización de información se ve limitada en la aplicación de soluciones y/o alternativas de aplicación. A continuación se presenta un listado de retos y desafíos que enfrenta la aplicación de este tipo de estrategias sobre repositorios digitales.

- El tiempo de desarrollo y adaptación de las técnicas de visualización planteadas de acuerdo a las características de recursos digitales dentro del repositorio. Esto supone agendar con cierta frecuencia y dependiendo de las versiones de la herramienta, el planteamiento de pruebas de usabilidad para facilitar el crecimiento y adaptación de las interfaces de acuerdo a criterios de acceso de mayor conveniencia para el desarrollo. Adicional a ello, esto implica considerar un rubro de dinero para los usuarios potenciales que puedan hacer las pruebas, y que de manera cercana puedan contar con la participación de expertos en el área de usabilidad para llevar a cabo un acompañamiento de la evolución que se realice sobre el desarrollo de la herramienta.
- La curva de aprendizaje por parte del equipo de desarrollo para llevar a cabo la implementación de técnicas de visualización. Es importante que el grupo de desarrollo encargado de llevar a cabo este tipo de soluciones, tenga nociones básicas para el manejo de estructuras y componentes que hacen parte de las bibliotecas de funciones de Vectores Gráficos Escalables (SVG), o APIs gráficas basadas en canvas soportadas por HTML5 y CSS3 según sea el caso. Tecnologías que soporten el uso de estructuras

radiales, hiperbólicas, en árbol, en categorías, y en general para el manejo de recursos gráficos en pantalla.

- Teniendo en cuenta el grado de complejidad del uso y aplicación de este tipo de estrategias, es recomendable que se conforme un grupo interdisciplinario de trabajo para que las responsabilidades de funcionamiento no solo recaigan en el equipo de desarrollo. Esto requiere de la participación de diseñadores o expertos en el manejo de componentes gráficos sobre la Web, expertos en usabilidad para identificar posibles fallas de uso de la herramienta y plantear sesiones de pruebas sobre las versiones beta que se puedan liberar dentro del proceso de desarrollo. De la misma forma contar con un equipo encargado de llevar a cabo la parte de manejo semántico o de vocabulario controlado (según sea el caso) para la gestión y administración del esquema de representación de conocimiento que se esté utilizando y se quiera adaptar en el repositorio, mediante el uso de estrategias de visualización de información. Finalmente disponer de un grupo de usuarios potenciales dependiendo del enfoque del repositorio digital (estudiantes, investigadores, profesionales), que utilicen este tipo de estrategias para llevar a cabo con ellos las sesiones de usabilidad y pruebas de las interfaces desarrolladas.
- El mantenimiento de este tipo de soluciones a largo plazo implica una tarea continua de adaptación y mantenimiento como es frecuente en la mayoría de proyectos informáticos. Por lo tanto, es indispensable que la arquitectura de desarrollo que se defina sea evidente en el manejo de las estructuras taxonómicas que se desplieguen, y tengan una estrecha relación con los esquemas de representación de conocimiento implementados en el repositorio. Para llevar a cabo esta actividad, debe existir una sinergia e interoperabilidad para la comunicación de cada una de las áreas y componentes que se definan en el framework. Esto facilitará una adaptación y enriquecimiento del repositorio, en la medida que se integren nuevos recursos digitales y se actualicen e integren de acuerdo a las nuevas áreas de conocimiento que se encuentren asociados cada uno de ellos.

Relacionados al esquema de representación de conocimiento podemos indicar algunas limitaciones en la capacidad de exploración de conceptos, de las cuales podemos mencionar:

- Terminologías asociadas con términos relacionados, o términos no preferidos en diferentes categorías jerárquicas. Dependiendo del tamaño, el número de conceptos por niveles de clasificación, y en especial, las relaciones del esquema de representación de conocimiento, es necesario realizar una adaptación profunda del algoritmo utilizado por la interfaz visual, con el propósito de desplegar de manera adecuada este tipo de relaciones y niveles de clasificación. Esto implica un esfuerzo adicional en la programación y adaptación dependiendo de la complejidad del esquema de representación de conocimiento.

- Esquemas de clasificación. Para usuarios finales a veces es complicado comprender el contexto de un concepto y las relaciones que vinculan los términos. Por lo tanto, es necesario llevar a cabo soportes o guías de aprendizaje en el marco visual para permitir que el usuario final tenga una mejor comprensión del sistema de representación del conocimiento que utilizando.
- Para mejorar el acceso de varios repositorios a través de la utilización de un esquema de representación del conocimiento es necesario llevar a cabo un mapeo completo de los conceptos y términos de todos los formatos SKOS. Si existen limitaciones para mapear estos vocabularios, las interfaces de búsqueda visual no podrían visualizar objetos de aprendizaje mediante el uso de repositorios. Estos problemas pueden afectar la usabilidad de la aplicación, si las interfaces visuales y los creadores de los repositorios no pueden proporcionar métodos de ayuda o guías para los participantes dentro de una estructura de navegación.

6.6 Resumen

En este capítulo se logró recopilar una serie de resultados asociados a estudios de usabilidad para la identificación de interfaces de búsqueda navegacionales, mediante el uso de técnicas de visualización. Esto facilitaría la ubicación de un término o temáticas de interés, a partir de la definición de un esquema de representación de conocimiento y del despliegue previo del número de recursos digitales asociados a cada término de consulta. Los análisis del estudio de usabilidad se llevaron a cabo a partir de resultados objetivos y subjetivos, mediante la aplicación de un caso de estudio para llevar a cabo una serie de actividades asociados con: procesos de búsqueda y localización de términos dentro de las interfaces desarrolladas. Adicional a ello, se llevó a cabo un estudio estadístico mediante análisis de correlación para analizar los resultados obtenidos y verificar la hipótesis propuesta, asociado con el uso de criterios de búsqueda basados en cobertura temática para facilitar el acceso a una colección de recursos digitales mediante técnicas de visualización. De manera paralela, se llevó a cabo un estudio de calidad de metadatos asociados con criterios de *completitud* y *exactitud*. Este estudio se planteó de manera complementaria, para identificar problemas asociados a la localización de recursos digitales a partir de la ausencia de metadatos, y adicionalmente para identificar criterios de búsqueda basados en un área de conocimiento.

A partir de los resultados obtenidos, se ha definido una serie de recomendaciones que facilitan: i) el desarrollo e implementación de estrategias de búsqueda visuales para facilitar el acceso a una colección de recursos digitales en un repositorio, y ii) la identificación de metadatos que permitan clasificar recursos digitales de acuerdo a una temática o área de conocimiento mediante el uso de un esquema de representación de conocimiento.

Pie de página:

[1] Disponible online en <http://www.ipl.org/> (último acceso 5 Mayo 2014)

Capítulo 7

Conclusiones y trabajo futuro

Este capítulo tiene como propósito presentar por un lado las conclusiones finales, producto de los resultados obtenidos a lo largo de este estudio. Por otro lado se exponen una serie de líneas de trabajo futuras que se lograron vislumbrar del estudio realizado. Finalmente se presentan una serie de recomendaciones que faciliten a creadores de repositorios digitales, la implementación de interfaces de búsquedas navegacionales a partir del uso de técnicas de visualización como alternativas de búsqueda que faciliten el acceso a una colección de recursos digitales.

Capítulo 7: Conclusiones y trabajo futuro

“Para las personas creyentes, Dios está al principio. Para los científicos está el final de todas sus reflexiones”. (Max Planck)

7.1 Resumen de objetivos y resultados

Este apartado tiene como finalidad presentar a modo resumen, todo los resultados que se han presentado en esta investigación. Para ello en la Tabla 23 se presentan los objetivos de partida planteados en el Capítulo 1 junto con los resultados obtenidos a lo largo del proceso de investigación realizado.

Tabla 23. Resumen de objetivos y resultados

Objetivo	Descripción	Resultado
(O1)	Estudiar el uso de esquemas de representación de conocimiento en repositorios digitales y las estrategias de búsquedas asociadas.	A
(O2)	Investigar e identificar las diferentes técnicas de visualización de información que permitan desplegar esquemas de representación de conocimiento en repositorios digitales.	B
(O3)	Desarrollar un prototipo que permita integrar técnicas de visualización y búsqueda de objetos de aprendizaje en repositorios digitales a partir de criterios definidos por sus metadatos.	C
(O3.1)	Diseñar y/o adaptar interfaces mediante técnicas de visualización a partir de esquemas de representación de conocimiento para desplegar información asociada a un área de conocimiento.	C1
(O3.2)	Desarrollar una herramienta que permita realizar extracción de recursos sobre un repositorio digital a partir de un área de conocimiento definido por un esquema de representación de conocimiento.	C2
(O3.3)	Analizar la calidad de los metadatos de los recursos digitales extraídos en (O3.2) e identificar criterios de búsquedas complementarios sobre cada una de las interfaces gráficas bajo estudio.	C3
(O4)	Realizar un estudio de usabilidad de las interfaces visuales diseñadas para identificar las ventajas y deficiencias de cada una de las técnicas de visualización bajo estudio.	D

Los resultados obtenidos de este proceso de investigación se llevaron a cabo en las siguientes actividades:

- A. Resultados que se abordaron en dos capítulos de la memoria: i) Capítulo 2 asociado con el estado de la cuestión, específicamente las secciones 2.1 y 2.2, y ii) Capítulo 5, Sección 5.5.1.

- B. Estudio realizado en las secciones 2.5 y 5.5.2, donde se identificaron las diferentes técnicas de visualización de información aptas para implementar sobre estrategias de búsquedas en repositorios digitales a partir de esquemas de representación de conocimiento.
- C. Actividad que se realizó a lo largo de los capítulos 4 y 5 mediante la definición de un modelo de trabajo que facilitó el desarrollo de un framework para el acceso a una colección de recursos digitales mediante el despliegue y la identificación de actividades para llevar a cabo el estudio.
 - C1. Actividad realizada en la Sección 4.3, mediante criterios de diseño y usabilidad relacionados a partir de Interacción Persona Ordenador (IPO).
 - C2. Proceso realizado en secciones 4.2.1 y 4.2.3, mediante el desarrollo de una herramienta Web Crawler para la extracción de recursos digitales de Europeana a partir de una rama de conocimiento definida por el tesauro AAT como esquema de representación de conocimiento.
 - C3. Evaluación presentada en secciones 4.2.4 y 5.3, a partir de un estudio de análisis de calidad de metadatos de recursos digitales de la biblioteca digital Europeana.
- D. Estudio realizado en secciones 5.1, 5.2 y 5.5, mediante análisis estadístico inferencial y análisis de correlación a partir de criterios de usabilidad previamente establecidos.

7.2 Conclusiones

En esta sección se describirán las conclusiones derivadas de la tesis, planteadas a partir de las áreas más representativas que se abordaron a lo largo del proceso de investigación.

7.2.1 Asociados al área de visualización de información

La visualización de información no deja de ser un área apasionante de gran interés sobre muchas disciplinas asociadas a la informática, la computación, y en general sobre todas aquellas líneas de trabajo que se encuentren involucradas con el tratamiento de grandes volúmenes de información. Por lo tanto, presenta una serie de retos y oportunidades orientadas a facilitar cada vez más la comprensión de información en sus diferentes formatos y representaciones, sobre una sociedad cada día con mayores demandas y necesidades para el acceso oportuno a información relevante, y con unas exigencias de calidad cada vez mayores. Por lo tanto, el campo de repositorios digitales no es ajeno a estas exigencias y necesidades y debe centrar sus esfuerzos en mejorar la accesibilidad de contenidos relevantes a partir de

criterios de búsqueda previamente establecidos, como por ejemplo una temática o área de conocimiento específico.

Es importante resaltar el papel de la visualización de información en el área de repositorios educativos para llevar a cabo actividades que faciliten: i) el acceso a un conjunto de recursos educativos por parte de usuarios convencionales (estudiantes y/o profesores); ii) tener una idea clara de los recursos digitales en todo su contexto para una mejor interpretación y asimilación de los mismos; iii) mayor comprensión de la temática o área de conocimiento objetivo de una consulta. Estos a su vez, son algunos de los desafíos de los actuales repositorios educativos (P. Davis & Connolly, 2007; A. Swan & Carr, 2008), que dentro del área de formación de un estudiante puede dificultar su proceso de aprendizaje, generando un progresivo abandono del uso de este tipo de herramientas.

7.2.2 Asociados a las técnicas de visualización

Una de las ventajas identificadas en las técnicas de visualización evaluadas, es que estas estrategias permitieron ofrecer un panorama general de las temáticas asociadas al contenido de toda la colección de recursos digitales. Cada técnica de visualización seleccionada facilitó la integración de estructuras taxonómicas a partir de la biblioteca de librerías gráficas trabajadas en el framework. Esta selección facilita su uso, y por lo tanto permite que cualquier creador de repositorios digitales pueda tomar decisiones para su selección desde el punto de vista: i) manejo de estructura jerárquica, ii) formas de representar datos, iii) manejo de relaciones jerárquicas, iv) facilidad de navegación, y otros criterios definidos en la Figura 51. Por lo tanto el uso de este tipo de técnicas de visualización podría ser un factor fundamental para vincularlos en las estrategias de búsqueda, dependiendo del perfil y de las preferencias de búsqueda que se establezcan dentro del repositorio, con el fin de incrementar el uso de repositorios y mejorar el tráfico de los mismos.

Es claro que las capacidades de percepción visual de las interfaces de búsqueda basadas en técnicas de visualización, son mayores que las planteadas por estrategias de búsqueda textual. De hecho uno de los factores que mayor perciben los usuarios dentro de una interfaz de búsqueda navegacional, es la capacidad de atención asociada a identificar las clasificaciones, jerarquías de navegación y cobertura temática, que se definen bajo una estructura taxonómica. De la misma forma, la percepción visual a partir de las formas, colores y tamaños que se presentaron en algunas interfaces visuales, favorecieron el uso de las estructuras de navegación, siendo otro de los factores valorados de manera positiva por la mayoría de participantes que hicieron parte de las pruebas de usabilidad. Por lo tanto fueron factores que influyeron notoriamente sobre el uso de estructuras de navegación, pero que no fueron abordadas en esta investigación de manera exhaustiva, siendo por tanto tema de estudios futuros mediante el uso de técnicas especializadas como *eye-tracking* (Rosch & Vogel-Walcutt, 2013).

7.2.3 Asociados a los esquemas de representación de conocimiento

En general, de acuerdo con los resultados de este estudio se considera que el desarrollo de interfaces gráficas asociados a esquemas de representación de conocimiento por parte de los creadores de repositorios, indican que estos deberían centrar sus esfuerzos en: i) definir niveles de relación y asociación entre recursos digitales, mediante el uso de estructuras de representación de conocimiento que soporten taxonomías (tesauros u ontología); ii) dependiendo de los niveles de jerarquía o profundidad que proporcione un repositorio, tratar de trabajar con interfaces fáciles de usar: (*tree*, *radial* o *icicle*) que soporten el uso de rutas de navegación para identificar el nivel de clasificación dentro de la estructura taxonómica; iii) vinculación de campos de metadatos relevantes (campos que permiten clasificar los recursos) con la estructura de navegación, para facilitar procesos de exploración por categorías o áreas de conocimiento. Esto es significativo para ofrecer a los usuarios alternativas de acceso más efectivas a una colección de recursos digitales; iv) integración visual en la estructura taxonómica y de manera proporcional, del número de recursos digitales que dispone cada categoría. Dado que la cobertura temática es un factor que mejoraría notablemente el acceso a una colección de recursos digitales, siempre y cuando los resultados desplegados sean visualmente estéticos para el participante. Esta condición es uno de los principios de usabilidad, que debe contener toda interfaz de búsqueda a nivel de percepción visual.

Por otro lado, los creadores de repositorios deberían considerar el diseño de interfaces de búsqueda visual a partir de esquemas de representación de conocimiento basados en: i) perfiles de usuario (estudiantes secundarios, estudiantes universitarios, profesionales o investigadores). Esta clasificación podría mejorar el aprendizaje de los usuarios finales que buscan entender las complejas relaciones de acuerdo con el término/concepto seleccionado; ii) seleccionar esquemas de representación de conocimiento basados en criterios como la clasificación de la taxonomía, los niveles de profundidad, niveles de asociación y número de términos relacionados por concepto seleccionado; iii) vinculación de recursos digitales de otros repositorios a partir de las asignaciones de términos y conceptos relacionados con otros esquemas de representación de conocimiento (ontologías, tesauros). Esta última condición podría mejorar el tiempo de respuesta de las consultas realizadas en los diferentes niveles de clasificación de términos / conceptos.

7.2.4 Asociados a los estudios de usabilidad

Los resultados del estudio de usabilidad aplicado sobre las interfaces implementadas, demostraron que la percepción visual de un participante (para determinar la cobertura temática a partir de un área de conocimiento específico) influye de manera significativa sobre la selección de una colección de recursos digitales. Estos resultados permiten determinar que un grupo de usuarios puede tener una mayor capacidad de decisión al momento de acceder sobre una colección de recursos digitales en función de: i) la cobertura temática del dominio de conocimiento a ser consultado, ii) la proporción de recursos digitales asociados a un dominio de

conocimiento y, finalmente, iii) la ubicación de recursos digitales en función de la estructura de navegación y clasificación taxonómica.

Adicionalmente dentro de los resultados obtenidos en las pruebas de usabilidad (Capítulo 5), para llevar a cabo la implementación de una interfaz usable es necesario conocer las condiciones, necesidades y propósitos del repositorio (descritos en profundidad en Sección 5.5.5). Por ejemplo, cuando el repositorio posee una gran profundidad dentro de su jerarquía taxonómica y un buen nivel de cobertura temática, las interfaces más efectivas son *icicle* y *tree*, aunque estéticamente no fueron las mejores en la valoración realizada por la mayoría de los participantes. Sin embargo, la interfaz *radial-search* presentó buenos resultados asociados a la *cobertura temática* y la *estética*, pero con bajos resultados a nivel de *clasificación*, por lo tanto, puede ser una interfaz apta para aplicar sobre repositorios que no presenten grandes niveles de clasificación definidos en sus jerarquías taxonómicas.

7.2.5 Asociados a los estudios de calidad de metadatos

El problema de la baja calidad asociado a la precisión de metadatos confirma los resultados de otras investigaciones relacionadas. Por ejemplo, Barton et al. (2003) expone algunas áreas donde encontró problemas asociados con algunos metadatos como: 1) la ortografía y las abreviaturas; 2) campos asociados con el proveedor de contenido; 3) campo de autor, 4) campos asociados con la temática, y 5) la descripción de la fecha. Estos mismos problemas se evidenciaron en la variedad de resultados encontrados en el estudio de completitud de metadatos examinados en Europeana (Sección 4.2.3 y 5.3), específicamente en los campos asociados con *title*, *subject*, *relation*, *type* y *language*. Uno de los errores más comunes que se hicieron evidentes a través del análisis fue la repetición del campo *título* para un gran número de recursos digitales. Por último, otro error común es que la descripción de los recursos no correspondía con el lenguaje definido originalmente por el recurso digital dentro de los elementos de metadatos *subject* y *type*.

Como conclusión de todo este estudio realizado de análisis de calidad de metadatos asociados a la rama de conocimiento de "*styles and periods*" de AAT, y en general sobre cualquier área de conocimiento, es importante definir metadatos que permitan clasificar el recurso digital a partir de una temática o área de conocimiento. Ofrecer un conjunto limitado de metadatos en las descripciones de recursos digitales dentro de un repositorio digital, generaría una gran variedad de resultados no relevantes dentro de procesos de búsqueda. La definición de estos metadatos, y la calidad de los mismos, ofrecería una variedad de criterios de búsqueda (temática, idioma, tipo de recurso, derechos de autor, etc.) que se pueden integrar dentro del desarrollo de interfaces visuales a partir de técnicas de visualización.

7.3 Recomendaciones

En esta sección presentamos a modo resumen algunas de las recomendaciones por áreas de interés que se han recopilado mediante una serie de criterios de valoración expuestos en la Sección 5.5.

7.3.1 Asociados a repositorios digitales

Sin lugar a dudas los repositorios digitales deben trabajar de la mano con estrategias que faciliten la interoperabilidad y reutilización de recursos digitales a partir del enriquecimiento semántico definido en sus metadatos. Sin embargo, se deben enfocar mayores esfuerzos para facilitar el acceso a grandes colecciones de recursos digitales mediante el planteamiento de estrategias para la gestión y mantenimiento de recursos digitales a partir de buenas prácticas de diseño y desarrollo. A continuación se extraen algunas de las recomendaciones más representativas asociadas al área de repositorios digitales presentadas en la Sección 5.5.1:

- *Gestión de calidad de metadatos*: Definición de políticas para la gestión y administración de la calidad de metadatos, que mejore las condiciones de precisión en la definición de metadatos, con el propósito de mejorar los resultados de búsqueda de recursos digitales relevantes.
- *Vinculación de campos de metadatos asociados por temáticas*: Definir estrategias mediante la implementación de metadatos que faciliten la búsqueda de recursos digitales a partir de una temática o área de conocimiento. Esta actividad se puede llevar a cabo mediante la definición de metadatos LOM como "*classification*" y "*relation*". Estos metadatos proporcionarían una clasificación por áreas o temas, que facilitaría los resultados de búsqueda de recursos digitales de acuerdo a un área de conocimiento específica con el propósito de simplificar el acceso sobre una colección de recursos digitales.
- *Esquema de representación de conocimiento*: Selección de un esquema de representación de conocimiento que disponga por lo menos de:
 - *Jerarquía taxonómica*: Una definición básica de su jerarquía taxonómica, con el propósito de facilitar la clasificación y uso de la técnica de visualización.
 - *Relaciones de jerarquía y asociación*: Se establezca una relación de asociación o jerarquía dentro de las estructuras de la estructura taxonómica definida, para favorecer la vinculación de términos afines (genéricos o específicos) al proceso de búsqueda.
 - *Manejo de vocabulario enriquecido*: Proporcionar capacidades de expresión dentro de las relaciones definidas a partir de términos complementarios que faciliten alternativas de búsqueda basadas en sinónimos, términos no preferidos, etc.

- *Definición de criterios de búsqueda:* Dependiendo de las estrategias de uso del repositorio, definir criterios de búsqueda a partir de la definición de metadatos basados en temáticas o áreas de conocimiento que faciliten la exploración e interacción de los usuarios sobre una colección de recursos digitales. De la misma forma mediante la integración de criterios definidos por el contexto educativo, perfiles de usuario, idioma y tipos de recursos digitales, etc.
- *Pruebas de usabilidad:* Definición y establecimiento de pruebas de usabilidad previos al desarrollo de una solución, que midan tanto la percepción como la interacción de los usuarios del repositorio, con el propósito de facilitar la evolución del diseño y funcionalidades de las interfaces para mejorar los procesos de búsqueda y acceso a recursos digitales relevantes. De igual manera, se recomienda realizar una programación periódica de estas pruebas una vez se encuentren las interfaces operativas, este proceso enriquecerá de manera significativa el uso y adecuación de estrategias complementarias para mejorar el acceso a recursos digitales dentro del repositorio.

7.3.2 Asociados al diseño de las interfaces

No todas las técnicas de visualización facilitan un proceso de búsqueda de recursos digitales, sin embargo ofrecen una serie de funcionalidades que mejoran el acceso y la satisfacción del usuario dentro de un repositorio. Por lo tanto, la selección de técnicas apropiadas va más allá de las funcionalidades que puede ofrecer, dado que la mayoría de ellas tiene como propósito ofrecer alternativas de exploración mediante una estructura taxonómica. Sin embargo, es necesario identificar una serie de factores técnicos asociados con las herramientas de desarrollo y bibliotecas de funciones que facilite la implementación de cada una de ellas en las interfaces (secciones 5.5.2 y 5.5.3). A continuación se destacan las siguientes recomendaciones:

- *Selección de bibliotecas de funciones gráficas:* Conocer las capacidades computacionales a nivel gráfico que pueda ofrecer la librería seleccionada para determinar las técnicas de visualización que soporte, y bibliotecas de funciones gráficas que permitan la integración de complementos de interacción que faciliten su uso.
- *Capacidad de respuesta:* Realizar pruebas para medir el tiempo de respuesta y rendimiento que ofrezca la técnica de visualización y librería seleccionada, sobre esquema de representación de conocimiento definido en el repositorio. Esto permitirá medir el tiempo de respuesta de navegación al cargar categorías y nodos asociados a cada nivel de profundidad dependiendo de la complejidad de la estructura taxonómica.
- *Integración y soporte de formatos:* Seleccionar bibliotecas de funciones con soportes de formatos de datos (JSON, XML, RDF) para el desarrollo e implementación de la estructura taxonómica, de acuerdo al esquema de representación de conocimiento

definido en el repositorio. Actualmente la mayoría de bibliotecas de funciones tienen un gran soporte para la implementación de estructuras taxonómicas basadas en formatos JSON, sin embargo estas estructuras taxonómicas se pueden integrar con recursos digitales para realizar consultas sobre formatos en XML, RDF o SKOS.

- *Escalabilidad y entorno de trabajo*: Ofrecida por las herramientas para adaptarse a cambios y actualizaciones que sufra el esquema de representación de conocimiento a partir de nuevos recursos digitales agregados al repositorio.
- *Rutas de navegación*: La definición de rutas de navegación apropiadas para guiar el proceso de exploración dentro de la estructura de navegación y la localización de recursos digitales asociados a temáticas afines.
- *Integración búsquedas complementarias*: La integración de estrategias de búsquedas complementarias a partir de las características de acceso que se definan en los recursos digitales sobre su contexto educativo (Educación media, avanzada, continuada, universitaria, profesional), perfiles de usuarios (maestros, estudiantes, profesionales), formatos del recurso digital (imagen, video, documento rtf, presentación, filminas, etc.), idioma del recurso digital.
- *Despliegue de recursos digitales*: Agregar estrategias de despliegue de recursos digitales relevantes dentro de un mecanismo de paginación, a partir de la valoración de recursos digitales realizadas por usuarios, de acuerdo a su perfil definido dentro del repositorio mediante evaluación realizada por: expertos, usuarios registrados, profesionales, estudiantes, etc.

7.4 Aportaciones originales

Una de las aportaciones centrales de esta investigación es ofrecer una alternativa de búsqueda visual a partir de técnicas de visualización para el acceso a una colección de recursos digitales sobre repositorios académicos. El propósito fue plantear una serie de recomendaciones para los creadores de repositorios digitales, antes de llevar a cabo un desarrollo formal de interfaces de búsqueda mediante el uso de técnicas de visualización. Esto disminuirá, previsiblemente el tiempo, esfuerzo y factores económicos en la implementación de este tipo de alternativas dentro de repositorios digitales. Entre las aportaciones complementarias que han surgido a partir de esta contribución principal, se pueden destacar las siguientes:

1. Un análisis de calidad de metadatos de recursos digitales de biblioteca digital Europea, donde se han identificado algunas de las limitaciones y se plantearon recomendaciones sobre el modelo de representación de datos para la búsqueda de recursos digitales a partir de la cobertura temática.

2. La aplicación de principios de Interacción Persona Ordenador (IPO) para plantear análisis de usabilidad de las interfaces diseñadas a partir de las técnicas de visualización.
3. La definición de una serie de recomendaciones para el desarrollo de interfaces visuales a partir de técnicas de visualización sobre repositorios digitales para el acceso a grandes volúmenes de recursos digitales.
4. El desarrollo de un framework e integración de métodos de búsquedas tradicionales, con métodos de búsqueda navegacionales a partir de técnicas de visualización.
5. La integración de métodos de valoración de recursos digitales dentro de resultados de búsqueda. Método complementario para desplegar el listado de recursos digitales relevantes a partir de una colección de recursos digitales identificados por área de conocimiento.

7.5 Trabajos futuros

A continuación se enumeran las líneas de trabajo futuro que se consideran importantes para continuar con esta investigación:

1. Una de las líneas o trabajos futuros dentro del área de la visualización de información en repositorios digitales, es el desarrollo de aplicaciones que faciliten el análisis mediante de informes estadísticos y toma de decisión para el análisis de grandes volúmenes de datos, orientado al área de analítica visual. Este tipo de soluciones facilitaría el planteamiento de estrategias orientadas a la preservación y divulgación de un repositorio, mediante el estudio de diversas áreas asociadas con:
 - El uso de objetos de aprendizaje (recursos relevantes, áreas de conocimiento mayor consultados, formatos de recursos mayor frecuentados, tipos de recurso de mayor preferencia, etc.).
 - Acceso de usuarios (perfiles de usuario, horas de mayor frecuencia de acceso, lugar de acceso, etc.), con el propósito de ofrecer alternativas que faciliten aumentar el tráfico para el uso de repositorios.
 - Identificar los términos más frecuente usados dentro de los procesos de consulta realizados por un usuario, con el propósito de seguir afinando las estrategias de búsqueda visual y ofrecer mejores alternativas de acceso.
2. Otra posible línea se puede orientar hacia el desarrollo de herramientas de transformación de datos, a partir del lenguaje definido por el esquema de representación de conocimiento de un repositorio (RDF, OWL). Una de las limitaciones que poseen la

mayoría de desarrollos basados en esquemas de representación de conocimiento es la compatibilidad y definición de formatos que permitan realizar actividades de transformación de datos soportados por cada una de las técnicas de visualización mediante formatos tipo JSON. Esto facilitaría el diseño e implementación de estructuras de navegación taxonómicas dentro de las interfaces de búsqueda navegacionales seleccionadas a partir de una técnica de visualización.

3. Una posible área de trabajo es la definición de interfaces de búsqueda navegacionales basados en esquemas de representación de conocimiento más complejos, a través del uso de ontologías y despliegue de resultados con mayores capacidades de expresión semánticas, a partir de la vinculación de recursos digitales mediante Linked Data. Tecnologías que favorecen una búsqueda más acertada de acuerdo a su expresividad semántica que favorece el uso de relaciones de asociación y de jerarquía, pero que por su grado de complejidad y tiempos de respuesta serían indispensables evaluarlas a la luz de la vinculación de resultados sobre interfaces de búsquedas navegacionales a partir de técnicas de visualización.
4. Otra posible línea sería la definición de una herramienta que permita la integración de resultados de búsqueda de repositorios digitales y las despliegue a través de interfaces de búsqueda navegacionales a partir de criterios definidos por áreas de conocimiento, tipo de recursos digitales, idioma, proveedor de contenidos, etc. Este tipo de soluciones podrían favorecer el análisis y definición de estrategias que faciliten el acceso a una colección de recursos digitales relevantes a partir de criterios de valoración realizados por perfiles de usuarios de acuerdo a la comunidad de usuarios de un repositorio definidos por (contexto académico, red social, perfil de usuario, etc.).
5. Finalmente asociado al área de usabilidad, se puede relacionar con el desarrollo de una herramienta que permita evaluar y seleccionar una técnica de visualización a partir de criterios asociados con: i) complejidad de la estructura taxonómica, ii) facilidad de uso, ii) navegación, iii) tiempo de respuesta de términos vinculados sobre diferentes niveles de profundidad taxonómica, iv) estética, y v) cobertura temática. Este tipo de estudios facilitaría el desarrollo de interfaces de búsqueda navegacionales basados en técnicas de visualización para su posterior implementación sobre repositorios digitales.

7.6 Publicaciones derivadas de la investigación

A lo largo del desarrollo de esta investigación se han realizado una serie de publicaciones en diversos ámbitos de divulgación.

7.6.1 Revistas internacionales

- 1) Gaona, P., Martin, D. Feroso, A., Sánchez, S. (2014) A usability study of taxonomy-visualization user interfaces in digital repositories. Journal Online Information review. Volume 38, Special Issue 2. Factor de impacto: 0,932
- 2) Gaona, P., Feroso, A., Sánchez, S. (en progreso). Exploring the coverage of the Europeana digital library using the AAT Thesaurus as a use case. Volume, Issue, pp.. Factor de Impacto:
- 3) Gaona, P., Stoitsis, G., Sánchez, S., Biniari, K. (en progreso). Visual framework for the access of several repositories through a knowledge representation scheme: Study case of Organic.Edunet and VOA3R/AGRIS.

7.6.2 Congresos

- 1) Martin, D., Sánchez, S., Gaona, P. & Marianos, N. (2013). Applying visualization techniques to develop interfaces for educational repositories: the case of Organic.Lingua and VOA3R. *Proceedings of the Learning Innovations and Quality: The Future of Digital Resources*, 60-67.
- 2) Gaona, P., Sánchez, S., Feroso, A. (2012). Análisis de cobertura del tesoro AAT en la biblioteca digital Europeana: ideas preliminares para su uso en Educación. Multidisciplinary symposium on the design and evaluation of digital content for education. SPDECE

7.6.3 Revistas nacionales indexadas

- 1) Gaona, P., Sánchez, S. (2014) Information Visualization: A proposal to improve search and access digital resources in repositories. Vol 34, No 1. Revista Ingeniería e investigación. ISSN: 2248-8723. Factor de Impacto: 0,034
- 2) Gaona, P., Sánchez, S., Feroso, A. (en impresión) Visualization techniques trough search interfaces in learning object repositories. Vol 18, No 41. Revista TECNURA. ISSN: 0123-921X.
- 3) Gaona, P., Sánchez, S. (2013). Modelo Informático para extracción de recursos digitales en Internet. Vol 17, Edición Especial. Revista TECNURA. ISSN: 0123-921X.

Anexo A

Estrategia de extracción de recursos digitales (normal vs refinada)

Tabla 24. Estrategia de extracción recursos digitales (Normal VS Refinada)

Exploración de términos AAT normal	Número de recursos digitales	Exploración de términos refinada	Número de recursos digitales
African	6913	<i>African styles and periods</i>	186
Rustic	669	<i>Rustic style</i>	68
Asian	2529	<i>Asian styles and periods</i>	189
prehistoric	2529	<i>prehistoric period</i>	114
Paleolithic	92	<i>Paleolithic period</i>	12
Lower Paleolithic	21	<i>"Lower Paleolithic" period</i>	10
Middle Paleolithic	14	<i>"Middle Paleolithic" period</i>	11
Upper Paleolithic	31	<i>"Upper Paleolithic" period</i>	26
Mesolithic	1453	<i>Mesolithic period</i>	104
Neolithic	3223	<i>Neolithic period</i>	348
Early Neolithic	2867	<i>"Early Neolithic"</i>	380
Middle Neolithic	83	<i>"Middle Neolithic" period</i>	37
Late Neolithic	702	<i>"Late Neolithic" period</i>	460
Chalcolithic	44	<i>Chalcolithic period</i>	38
Bronze Age	9274	<i>"Bronze Age" period</i>	5833
Iron Age	20331	<i>"Iron Age" period</i>	11222
protohistoric	2	<i>protohistoric period</i>	0
European	19257	<i>european styles and periods</i>	390
Netherlandish	582	<i>Netherlandish culture regions</i>	1
Magic Realist	4	Magic Realist	6
armorial	522	<i>armorial ceramical</i>	9
Art Deco	485	<i>"Art Deco" style art</i>	150
Art Nouveau	3214	<i>"Art Nouveau" style art</i>	552
Beaux-Arts	13055	<i>"Beaux-Arts" styles</i>	394
Biedermeier	519	Biedermeier	519
Concrete art	420	<i>"Concrete art"</i>	301
Baroque Revival	144	Baroque Revival	146

Byzantine Revival	4	Byzantine Revival	4
Egyptian Revival	5	Egyptian Revival	5
Etruscan Style	7	Etruscan Style	8
Gothic Revival	626	<i>"Gothic Revival"</i>	406
Greek Revival	32	Greek Revival	43
Pompeian Style	2	Pompeian Style	2
Renaissance Revival	117	Renaissance Revival	124
Rococo Revival	11	Rococo Revival	14
Romanesque Revival	19	Romanesque Revival	19
International Style	7803	<i>International Style architecture</i>	7337
Modernist	254	<i>Modernist art</i>	125
Neo-Romantic	10	Neo-Romantic	13
Neoclassical	1766	<i>Neoclassical art</i>	195
Abbotsford	25	Abbotsford	162
Japonaiserie	6	Japonaiserie	8
Cubist	39	Cubist	80
Dada	3014	<i>Dada art</i>	944
Expressionist	42	Expressionist	69
Impressionist	149	Impressionist	436
Post-Impressionist	11	Post-Impressionist	19
Surrealist	888	<i>Surrealist art</i>	845
Symbolist	53	Symbolist	106
Oceanic	14180	<i>ocean styles and periods</i>	30
ancient	17618	<i>ancient art</i>	58
Dark Ages	1460	<i>Dark Ages periods</i>	24
Stone Age	1274	<i>Stone Age period</i>	59
Epipaleolithic	9	<i>Epipaleolithic period</i>	9
Early Chalcolithic	6	<i>Early Chalcolithic period</i>	0
Middle Chalcolithic	3	<i>Middle Chalcolithic period</i>	1
Late Chalcolithic	10	<i>Late Chalcolithic period</i>	0
Eneolithic	7	<i>Eneolithic period</i>	5
Final Neolithic	23	<i>Final Neolithic period</i>	7
Abstract	11518	<i>Abstract art</i>	2981
Classical Revival	77	Classical Revival	88
Albanian	130	<i>Albanian culture regions</i>	1
Austrian	2844	<i>Austrian culture regions</i>	9
Belgian	8618	<i>Belgian culture regions</i>	0
British	74608	<i>British culture regions</i>	93
Bulgarian	1038	<i>Bulgarian culture regions</i>	5
Czech	33511	<i>Czech culture regions</i>	46

Scandinavian	2632	<i>Scandinavian culture regions</i>	10
Danish	7057	<i>Danish culture regions</i>	2
Dutch	36364	<i>Dutch culture regions</i>	11
English	12791	<i>English culture regions</i>	49
Finnish	0	<i>Finnish culture regions</i>	8
French	0	<i>French culture regions</i>	44
German	0	<i>German culture regions</i>	121
Hungarian	0	<i>Hungarian culture regions</i>	22
Italian	42135	<i>Italian culture regions</i>	122
Norwegian	5059	<i>Norwegian culture regions</i>	16
Polish	34488	<i>Polish culture regions</i>	636
Portuguese	243	<i>Portuguese culture regions</i>	4
Romanian	3052	<i>Romanian culture regions</i>	1
Scottish	6568	<i>Scottish culture regions</i>	1
Spanish	40192	<i>Spanish culture regions</i>	1677
Swiss	22590	<i>Swiss culture regions</i>	5
Welsh	8261	<i>Welsh culture regions</i>	0
Yugoslav	693	<i>Yugoslav culture regions</i>	4
Irish	6676	<i>Irish culture regions</i>	4
Russian	8709	<i>Russian culture regions</i>	28
Neues Bauen	0	Neues Bauen	122
Pictorialist	0	Pictorialist	146
Secession Movement	0	Secession Movement	6
Modern Movement	0	Modern Movement	290
Unanimist	0	Unanimist	2
Neo-Florentine	0	Neo-Florentine	0
Rosicrucian	45	Rosicrucian	47
Straight	0	<i>Straight art</i>	481
Realist	0	<i>Realist art</i>	247
Romantic	0	<i>Romantic art</i>	1739
Moorish Revival	0	Moorish Revival	0
Wemyss	0	<i>Wemyss ceramical</i>	29
New Vision	0	<i>New Vision art</i>	100
Serbian	0	<i>Serbian culture regions</i>	14
Slovene	0	<i>Slovene culture regions</i>	1
Croatian	0	<i>Croatian culture regions</i>	3
Wendish	0	<i>Wendish culture regions</i>	0
Pre-Conquest	0	<i>Pre-Conquest culture regions</i>	0
Nordic	0	<i>Nordic culture regions</i>	8
Belorussian	0	<i>Belorussian culture regions</i>	0

Barbizon School	128	Barbizon School	119
Nèò-Réalisme	0	neo realism	0
Modern	48030	<i>Modern art</i>	1374
Contemporary	2964	<i>Contemporary art</i>	800
Greek	0	<i>Greek culture regions</i>	21
Neo-Idealism	0	Neo-Idealism	0
Rockingham ware	0	Rockingham ware	0
Anglo-Japanese	1	Anglo-Japanese	1
neo-Russian Style	3	<i>neo-Russian Style</i>	3
Nazarene	75	Nazarene	76
Total	559.552	TOTAL	44.280

Anexo B

Cuestionario de datos prueba de usabilidad

USER ID	
PROFILE QUESTIONNAIRE	

This questionnaire helps us to define your user profile, please be sincere in your answers. No personal information will be stored, remember we are testing the software, not you.




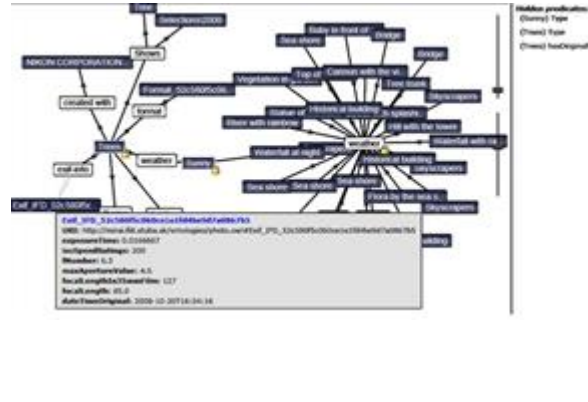
Are you familiar with these terms?

() Metadata () Semantic () Thesaurus () Ontology

Profession: _____

Working at: _____

What do you know about this searching interfaces?

	<p>Textual Search</p> <p>Never used ()</p> <p>I've used it, but I don't know when using this searching method is more useful than using others. ()</p> <p>I've used it, know how it Works and when to use it instead other searching methods. ()</p>
	<p>Directory / Categories Browsing</p> <p>Never used ()</p> <p>I've used it, but I don't know when using this searching method is more useful than using others. ()</p> <p>I've used it, know how it Works and when to use it instead other searching methods. ()</p>
	<p>Tags Search</p> <p>Never used ()</p> <p>I've used it, but I don't know when using this searching method is more useful than using others. ()</p> <p>I've used it, know how it Works and when to use it instead other searching methods. ()</p>
	<p>Semantic Search</p> <p>Never used ()</p> <p>I've used it, but I don't know when using this searching method is more useful than using others. ()</p> <p>I've used it, know how it Works and when to use it instead other searching methods. ()</p>

Please provide us an email in case we've to contact you for further information about this trial:

EMAIL: _____

Referencias Bibliográficas

- Aitchison, J., Gilchrist, A., & Bawden, D. (2001). *Thesaurus construction and use: a practical manual* (Vol. 7): Fitzroy Dearborn.
- Aitta, M., Kaleva, S., & Kortelainen, T. (2008). Heuristic evaluation applied to library web services. *New Library World*, 109(1/2), 25-45.
- Alva, M., Martinez, A., Suarez, M., Labra, J., Cueva, J., & Sagastegui, H. (2010). Towards the evaluation of usability in educative websites. *International Journal of Technology Enhanced Learning*, 2(1), 145-161.
- Allan, J., Leuski, A., Swan, R., & Byrd, D. (2001). Evaluating combinations of ranked lists and visualizations of inter-document similarity. *Information processing & management*, 37(3), 435-458.
- Allen, R., Obry, P., & Littman, M. (1993). *An interface for navigating clustered document sets returned by queries*. Paper presented at the Proceedings of the conference on Organizational computing systems, Milpitas, California, United States.
- ANSI/NISO. (2005). ANSI/NISO Z39.19 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.
- Arano, S., & Codina, L. (2004). La estructura conceptual de los tesauros en el entorno digital: ¿nuevas esperanzas para viejos problemas. *Jornades Catalanes d'Informació i Documentació*, 9, 14.
- Armstrong, C., Fenton, R., Lonsdale, R., Stoker, D., Thomas, R., & Urquhart, C. (2001). A study of the use of electronic information systems by higher education students in the UK. *Program: electronic library and information systems*, 35(3), 241-262.
- Aroyo, L., Dolog, P., Houben, G.-J., Kravcik, M., Naeve, A., Nilsson, M., & Wild, F. (2006). Interoperability in personalized adaptive learning. *Educational Technology & Society*, 9(2), 4-18.
- Aula, A., & Kåki, M. (2005). Less is more in Web search interfaces for older adults. *First Monday*, 10(7-4).
- Baeza, R., & Castillo, C. (2004). Crawling the infinite Web: five levels are enough. *Algorithms and Models for the Web-Graph*, 3243, 156-167.
- Baker, K. D. (2006). Learning objects and process interoperability. *International Journal on E-Learning*, 5(1), 167-172.
- Barton, J., Currier, S., & Hey, J. M. N. (2003). *Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice*. Paper presented at the Dublin Core Conference: Supporting Communities of Discourse and

- Practice - Metadata Research and Applications, September 28 - October 2 (2003), Seattle, Washington, USA.
- Battista, G., Eades, P., Tamassia, R., & Tollis, I. (1998). *Graph drawing: algorithms for the visualization of graphs*: Prentice Hall PTR.
- Bederson, B., Shneiderman, B., & Wattenberg, M. (2002). Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies. *ACM Transactions on Computer Graphic*, 21(4), 833-854.
- Benford, S., Snowdon, D., Greenhalgh, C., Ingram, R., Knox, I., & Brown, C. (1995). VR-VIBE: A Virtual Environment for Co-operative Information Retrieval. *Computer Graphics Forum*, 14(3), 349-360. doi: 10.1111/j.1467-8659.1995.cgf143_0349.x
- Bevan, N., & Macleod, M. (1994). Usability measurement in context. *Behaviour & Information Technology*, 13(1-2), 132-145.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22.
- Blanch, R., & Lecolinet, E. (2007). Browsing Zoomable Treemaps: Structure-Aware Multi-Scale Navigation Techniques. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1248-1253. doi: 10.1109/tvcg.2007.70540
- Blocks, D., Binding, C., Cunliffe, D., & Tudhope, D. (2002). Qualitative evaluation of thesaurus-based retrieval *Research and Advanced Technology for Digital Libraries* (pp. 346-361): Springer.
- Boldi, P., Codenotti, B., Santini, M., & Vigna, S. (2004). Ubicrawler: A scalable fully distributed web crawler. *Software: Practice and Experience*, 34(8), 711-726.
- Booth, P. (1989). *An introduction to human-computer interaction*. London, UK: Psychology Press.
- Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *Professional Communication, IEEE Transactions on*, 43(3), 261-278.
- Bostock, M., & Heer, J. (2009). Protovis: A graphical toolkit for visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6), 1121-1128.
- Boulos, M., Warren, J., Gong, J., & Yue, P. (2010). Web GIS in practice VIII: HTML5 and the canvas element for interactive online mapping. *International journal of health geographics*, 9(1), 14.
- Brachman, R., & Schmolze, J. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive science*, 9(2), 171-216.
- Bruce, T., & Hillmann, D. (2004). *The continuum of metadata quality: defining, expressing, exploiting*. Chicago, IL: D.Hillman & E.Westbrooks (Eds.).
- Bruno, D., & Richmond, H. (2003). The truth about taxonomies. *Information management journal*, 37(2), 48-52.
- Buchanan, S., & Salako, A. (2009). Evaluating the usability and usefulness of a digital library. *Library Review*, 58(9), 638-651.
- Bui, Y., & Park, J. (2006). An assessment of metadata quality: A case study of the national science digital library metadata repository. *H. Moukdad (ed.) Proceedings of CAIS/ACSI 2006 Information Science Revisited: Approaches to Innovation*, 13.
- Buttenfield, B. (1999). Usability evaluation of digital libraries. *Science & Technology Libraries*, 17(3-4), 39-59.
- Card, S., Mackinlay, J., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Castells, L. (2007). Los protocolos de pensamiento en voz alta como instrumento para analizar el proceso de escritura. *Revista española de lingüística aplicada*(20), 27-36.
- Castells, P. (2002). Aplicación de técnicas de la web semántica. *línea*. Universidad Autónoma de Madrid.
- Castillo, C. (2005). Effective web crawling. *ACM Transactions on Computer Graphic*, 39, 55-56.

- Castillo, L. (2006). Elaboración de un tesoro de información de actualidad y conversión en red semántica para su empleo en un sistema de recuperación periodístico.
- Cechinel, C., Sánchez, S., & Sicilia, M. (2009). *Empirical analysis of errors on human-generated learning objects metadata*.
- Cechinel, C., Silva, S., Ochoa, X., Sánchez, S., & Sicilia, M. (2012). *Populating Learning Object Repositories with Hidden Internal Quality Information*. Paper presented at the Proceedings of the 2nd Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2012). Manouselis, N., Draschler, H., Verber, K., and Santos, OC (Eds.). Published by CEUR Workshop Proceedings.
- Centelles, M. (2005). Taxonomías para la categorización y la organización de la información en sitios web.
- Cervera, J., López, M., Fernández, C., & Sánchez, S. (2009). Quality metrics in learning objects. *Metadata and Semantics*, 135-141.
- Cleveland, W. (1984). *Graphs in scientific publications*. 38. (4)
- Cleveland, W., Diaconis, P., & McGill, R. (1982). *Variables on scatterplots look more highly correlated when the scales are increased*: Defense Technical Information Center.
- Clyphan, R., & Sugimoto. (2009). Europeana Semantic Element ESE v3.4.1, available online at: <http://pro.europeana.eu/web/guest/technical-requirements/>, (Acceso: 2 Mayo, 2014).
- Cockburn, A., & McKenzie, B. (2000). An evaluation of cone trees *People and Computers XIV—Usability or Else!* (pp. 425-436): Springer.
- Codina, L. (2000). Evaluación de recursos digitales en línea: conceptos, indicadores y métodos. *Revista española de documentación científica*, 23(1), 9-44.
- Corcho, O., Fernández, M., & Gómez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data & knowledge engineering*, 46(1), 41-64.
- Crockford, D. (2006). *The application/json media type for JavaScript Object Notation (JSON)*.
- Cugini, J., Laskowski, S., & Sebrechts, M. (2000). *Presenting search results: design, visualization and evaluation*. Paper presented at the Proceedings of Information Doors. Where Information Search and Hypertext Link Workshop.
- Chen, C. (2004). *Information visualization: Beyond the horizon*: Springer.
- Chirita, P., Gavriloiu, R., Ghita, S., Nejdl, W., & Paiu, R. (2005). Activity based metadata for semantic desktop search *The Semantic Web: Research and Applications* (pp. 439-454): Springer.
- Chrysafiadi, K., & Virvou, M. (2013). A knowledge representation approach using fuzzy cognitive maps for better navigation support in an adaptive learning system. *SpringerPlus*, 2(1), 81.
- Chuanjun, S. (2004). On the Evaluation of the Quality of Digital Collections. *The Journal of The Library Science In China*, 4.
- Davis, P., & Connolly, M. (2007). Institutional repositories: evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*, 13(3/4).
- Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation? *AI magazine*, 14(1), 17.
- De la Prieta, F., & Gil, A. (2010). A multi-agent system that searches for learning objects in heterogeneous repositories. *Trends in Practical Applications of Agents and Multiagent Systems*, 355-362.
- Design, P. (1998). Visual Thesaurus: URL: <http://www.visualthesaurus.com> (Acceso: 2 de Mayo, 2014).
- Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., & Taibi, D. (2012). *Linked Education: interlinking educational Resources and the Web of Data*. Paper presented at the Proceedings of the 27th Annual ACM Symposium on Applied Computing.

- Dinkla, K., Westenberg, M., Timmerman, H., van Hijum, S., & van Wijk, J. (2011). Comparison of multiple weighted hierarchies: visual analytics for microbe community profiling. *Computer Graphics Forum*, 30, 1141-1150.
- Dix, A., Janet, E., Abowd, G., & Beale, R. (2003). *Human computer interaction*. NY: Prentice-Hall, Inc.
- Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., & van de Sompel, H. (2010). *The Europeana Data Model (EDM)*. Paper presented at the Proceedings of IFLA, Gothenburg, Sweden.
- Downes, S. (2007). Models for sustainable open educational resources. *Interdisciplinary Journal of Knowledge and Learning Objects*, 3.
- Draper, G., Livnat, Y., & Riesenfeld, R. (2009). A Survey of Radial Methods for Information Visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15(5), 759-776. doi: 10.1109/tvcg.2009.23
- Dumas, J. (1999). *A practical guide to usability testing*: Intellect Books.
- Duncan, S. (2009). *Patterns of learning object reuse in the Connexions repository. PhD Thesis*.
- Durkin, J. (1994). *Expert systems: design and development*: New York: Maxwell Macmillan International.
- Eades, P., Sugiyama, K. (1990). How to Draw a Directed Graph *J. Information Processing* (Vol. vol. 13 no. 4, pp. pp. 424-434).
- Edwards, J., McCurley, K., & Tomlin, J. (2001). *An adaptive model for optimizing performance of an incremental web crawler*.
- English, J., Hearst, M., Sinha, R., Swearingen, K., & Lee, K. (2002). Flexible search and navigation using faceted metadata: Technical report, University of Berkeley, School of Information Management and Systems, 2003. Submitted for publication.
- Fang, X., & Holsapple, C. W. (2007). An empirical study of web site navigation structures' impacts on web site usability. *Decision Support Systems*, 43(2), 476-491.
- Fekete, J. (2004). *The infovis toolkit*. Paper presented at the Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on.
- Fekete, J., Hémy, P., Baudel, T., & Wood, J. (2011). *Obvious: A meta-toolkit to encapsulate information visualization toolkits - one toolkit to bind them all*. Paper presented at the Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on.
- Fernández, A. (2011). La construcción de tesauros académicos: un modelo general y un método inductivo con aplicación al "e-learning". Tesis Doctoral.
- Fernández, J. (2001). *Modelos de Recuperación de Información basados en Redes de Creencia*. Tesis Doctoral Universidad de Granada.
- Flouris, G., Plexousakis, D., & Antoniou, G. (2003). Describing knowledge representation schemes: A formal account. *Technical Rep. No. TR-320, Institute of Computer Science Foundation for Research and Technology Hellas (ICS-FORTH), Crete, Greece*.
- Foltz, M. (1997). An information space design rationale.
- Fox, E., Gonçalves, M., & Kipp, N. (2002). Digital libraries *Handbook on Information Technologies for Education and Training* (pp. 623-641): Springer.
- Fox, E., Hix, D., Nowell, L., Brueni, D., Wake, W., Heath, L., & Rao, D. (1999). Users, user interfaces, and objects: Envision, a digital library. *Journal of the American Society for Information Science*, 44(8), 480-491.
- Fox, E., & Logan, E. (2005). An Asian digital libraries perspective. *Information processing & management*, 41(1), 1-4.
- Friesen, N. (2001). What are educational objects? *Interactive learning environments*, 9(3), 219-230.
- Friesen, N. (2009). Open educational resources: New possibilities for change and sustainability. *The International Review of Research in Open and Distance Learning*, 10(5).

- Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). *Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Fulton, S., & Fulton, J. (2013). *HTML5 Canvas*: O'Reilly Media.
- Furnas, G. (1986). *Generalized fisheye views* (Vol. 17): ACM.
- Furnas, G., & Zacks, J. (1994). *Multitrees: enriching and reusing hierarchical structure*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Gaona, P., Feroso, A., & Sánchez, S. (n.p.). Exploring the coverage of Europeana through knowledge representation scheme: AAT thesaurus. *The Electronic Library (in press)*.
- Gaona, P., Martín, D., Feroso, A., & Sánchez, S. (2014). A usability study of taxonomy-visualization user interfaces in digital repositories. *Journal Online Information review*, 38(2).
- Gaona, P., Sánchez, S., & Gaona, E. (2013). Prototipo informático para extracción de recursos digitales sobre internet. *Revista Tecnura*, 17, 79-92.
- Gaona, P., Sánchez, S., & Montenegro, C. (2014). Visualization of information: a proposal to improve the search and access to digital resources in repositories. *INGENIERÍA E INVESTIGACIÓN*, 34(1), 83-89.
- Gašević, D., Djurić, D., & Devedzic, V. (2009). *Model driven engineering and ontology development*: Springer-Verlag Berlin Heidelberg.
- Gašević, D., & Hatala, M. (2006). Ontology mappings to improve learning resource search. *British Journal of Educational Technology*, 37(3), 375-389.
- Genesereth, M., & Ketchpel, S. (1994). Software agents. *Commun. ACM*, 37(7), 48-53, 147.
- Gero, J. (1990). Design prototypes: a knowledge representation schema for design. *AI magazine*, 11(4), 26.
- Gilchrist, A. (2003). Thesauri, taxonomies and ontologies—an etymological note. *Journal of documentation*, 59(1), 7-18.
- Gilchrist, A., Kibby, P., Mahon, B., & Libra, T. (2000). *Taxonomies for business: access and connectivity in a wired world*: TFPL London.
- Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, 10(4), 289-309.
- Gómez, A. (1999). Evaluation of taxonomic knowledge in ontologies and knowledge bases.
- Gonçalves, M., Moreira, B., Fox, E., & Watson, L. (2007). What is a good digital library? – A quality model for digital libraries. *Information Processing & Management*, 43(5), 1416-1437. doi: 10.1016/j.ipm.2006.11.010
- González, V., & Kobsa, A. (2003). *Benefits of information visualization systems for administrative data analysts*. Paper presented at the Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on.
- Graham, M., & Kennedy, J. (2005). Extending taxonomic visualisation to incorporate synonymy and structural markers. *Information Visualization*, 4(3), 206-223.
- Graham, M., & Kennedy, J. (2010). A survey of multiple tree visualisation. *Information Visualization*, 9(4), 235-252.
- Graham, M., Kennedy, J., & Benyon, D. (2000). Towards a methodology for developing visualizations. 53. (5)
- Graham, M., Kennedy, J., & Benyon, D. (2000). Towards a methodology for developing visualizations. *International Journal of Human-Computer Studies*, 53(5), 789-807.
- Granollers, T. (2004). MPLu+ a. Una metodología que integra la Ingeniería del Software, la Interacción Persona-Ordenador y la Accesibilidad en el contexto de equipos de desarrollo multidisciplinares.
- Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220.

- Guarino, N., & Poli, R. (1995). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human Computer Studies*, 43(5), 625-640.
- Guha, R., McCool, R., & Miller, E. (2003). *Semantic search*. Paper presented at the Proceedings of the 12th international conference on World Wide Web.
- Harger, J., & Crossno, P. (2012). *Comparison of open-source visual analytics toolkits*. Paper presented at the Proceedings of the SPIE Conference on Visualization and Data Analysis.
- Hargittai, E. (2004). Classifying and coding online actions. *Social Science Computer Review*, 22(2), 210-227.
- Hartson, H., Shivakumar, P., & Pérez, M. (2004). *Usability inspection of digital libraries: a case study*. 4. (2)
- Haslhofer, B., & Isaac, A. (2011). *data. europeana. eu: The Europeana Linked Open Data Pilot*. Paper presented at the International Conference on Dublin Core and Metadata Applications, 21 - 23 September 2011, The Hague, Netherlands.
- Hassan, Y., Martín, F., & Iazza, G. (2004). Diseño Web centrado en el usuario: usabilidad y arquitectura de la información. *Hipertext. net*(2).
- Hayati, Z., & Jowkar, T. (2008). Adoption of electronic reference materials in academic libraries of Iran. *The International Information & Library Review*, 40(1), 52-63.
- Hearst, M. (1995). *TileBars: visualization of term distribution information in full text information access*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Hearst, M. (2009). *Search user interfaces*. New York, NY: Cambridge University Press.
- Hearst, M., & Karadi, C. (1997). *Cat-a-Cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy*. Paper presented at the ACM SIGIR Forum.
- Heer, J., & Bostock, M. (2010). *Crowdsourcing graphical perception: using mechanical turk to assess visualization design*.
- Heer, J., Card, S. K., & Landay, J. A. (2005). *Prefuse: a toolkit for interactive information visualization*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Heer, J., Kong, N., & Agrawala, M. (2009). *Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Hemmje, M., Kunkel, C., & Willett, A. (1994). *LyberWorld—a visualization user interface supporting fulltext retrieval*. Paper presented at the Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval.
- Herman, I., Melancon, G., & Marshall, M. S. (2000). Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 6(1), 24-43. doi: 10.1109/2945.841119
- Hetzler, E., & Turner, A. (2004). Analysis experiences using information visualization. *Computer Graphics and Applications, IEEE*, 24(5), 22-26.
- Hitchcock, S., Woukeu, A., Brody, T., Carr, L., Hall, W., & Hamad, S. (2003). *Evaluating Citebase, an open access Web-based citation-ranked search and impact discovery service*. Monograph (Technical Report) Southampton.
- Hodgins, W., & Conner, M. (2000). Everything you ever wanted to know about learning standards but were afraid to ask. *Learning in the New Economy e-Magazine (LiNE Zine)*.
- Hoffman, P., Grinstein, G., Marx, K., Grosse, I., & Stanley, E. (1997). *DNA visual and analytic data mining*. Paper presented at the Visualization'97., Proceedings.

- Holzschuher, F., & Peinl, R. (2013). *Performance of graph query languages: comparison of cypher, gremlin and native access in Neo4j*. Paper presented at the Proceedings of the Joint EDBT/ICDT 2013 Workshops.
- Howland, J., Jonassen, D., & Marra, R. (2012). *Meaningful learning with technology*. Pearson Upper Saddle River, NJ.
- Huang, M., Liang, J., & Nguyen, Q. (2008, 26-28 Aug. 2008). *A Usability Study on the Use of Multi-Context Visualization*. Paper presented at the CGIV '08. Fifth International Conference on Computer Graphics, Imaging and Visualisation, 2008. , Penang.
- Hughes, B. (2005). Metadata quality evaluation: Experience from the open language archives community. *Digital Libraries: International Collaboration and Cross-Fertilization*, 135-148.
- ISO. (1985). ISO 5964-1985. Guidelines for the establishment and development of multilingual thesauri.
- ISO. (1998). *ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 11: Guidance on Usability*.
- ISO. (2001). ISO/IEC 9126-1:2001 Software product evaluation - Quality characteristics and guidelines for their use. .
- Jacob, R., & Karn, K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3), 4.
- Jain, P., & Babbar, P. (2006). Digital libraries initiatives in India. *The International Information & Library Review*, 38(3), 161-169.
- Jeng, J. (2005). *Usability assessment of academic digital libraries: Effectiveness, efficiency, satisfaction, and learnability*. 55. (2-3)
- Jeng, J. (2009). What should we take into consideration when we talk about usability? *G. Tsakonas & C. Papatheodorou (Eds.) , Evaluation of digital libraries, Chandos Publishing, Oxford.*, 63-73.
- Jovanović, J., Gašević, D., Knight, C., & Richards, G. (2007). Ontologies for effective use of context in e-learning settings. *Educational Technology & Society*, 10(3), 47-59.
- Kani, E., Ghinea, G., & Chen, S. Y. (2006). Digital libraries: what do users want? *Online Information Review*, 30(4), 395-412.
- Keim, D., Mansmann, F., Schneidewind, J., & Schreck, T. (2006). *Monitoring network traffic with radial traffic analyzer*. Paper presented at the Visual Analytics Science And Technology, 2006 IEEE Symposium On.
- Kelty, C., Burrus, C., & Baraniuk, R. (2008). Peer review anew: Three principles and a case study in postpublication quality assurance. *Proceedings of the IEEE*, 96(6), 1000-1011.
- Khoo, M., Kusunoki, D., & MacDonald, C. (2012, 4-7 Jan. 2012). *Finding Problems: When Digital Library Users Act as Usability Evaluators*. Paper presented at the System Science (HICSS), 2012 45th Hawaii International Conference on.
- Kim, H., & Kim, Y. (2008). Usability study of digital institutional repositories. *Electronic Library, The*, 26(6), 863-881.
- Klerkx, J., Duval, E., & Meire, M. (2004). Using information visualization for accessing learning object repositories. 2004. IV 2004. *Proceedings. Eighth International Conference on Information Visualisation, London*, 465-470. doi: 10.1109/iv.2004.1320185
- Klerkx, J., Meire, M., Ternier, S., Verbert, K., & Duval, E. (2005). *Information visualisation: towards an extensible framework for accessing learning object repositories*. Paper presented at the Proceedings of World Conference on Educational Multimedia, Hypermedia & Telecommunications.
- Kobsa, A. (2004). *User experiments with tree visualization systems*. Paper presented at the Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on.

- Komlodi, A., Marchionini, G., & Soergel, D. (2007). Search history support for finding and using information: User interface design recommendations from a user study. *Information Processing & Management*, 43(1), 10-29. doi: 10.1016/j.ipm.2006.05.017
- Kruskal, J., & Landwehr, J. (1983). Icicle plots: Better displays for hierarchical clustering. *The American Statistician*, 37(2), 162-168.
- Kumar, V., Nesbit, J., & Han, K. (2005). *Rating learning object quality with distributed bayesian belief networks: The why and the how*. Paper presented at the Advanced Learning Technologies, 2005. ICALT 2005. Fifth IEEE International Conference on.
- Lagoze, C., Van de Sompel, H., Johnston, P., Nelson, M., Sanderson, R., & Warner, S. (2007). Open Archives Initiative Object Reuse and Exchange (OAI-ORE): Technical report, Open Archives Initiative.
- Lamping, J., Rao, R., & Pirolli, P. (1995). *A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Lamping, L., & Rao, R. (1996). The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierarchies. *Journal of Visual Languages and Computing*, 7(1), pp. 33-55.
- Lee, M., Reilly, R., & Butavicius, M. (2003). *An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data*. Paper presented at the ACM International Conference Proceeding Series.
- Lee, M., & Vickers, D. (1998). Psychological approaches to data visualisation: DTIC Document.
- Lei, M. (2008). Knowledge organization systems (KOS). *Knowledge Organization*, 35(2-3), 160-182.
- Lewis, C. (1982). *Using the "thinking-aloud" method in cognitive interface design*: IBM TJ Watson Research Center.
- Li, J., Gasevic, D., Nesbit, J., & Richards, G. (2005). *Ontology Mappings Enable Interoperation of Knowledge Domain Taxonomies*. Paper presented at the 2nd LORNET international annual conference.
- Lidwell, W., Holden, K., & Butler, J. (2010). *Universal Principles of Design: 125 Ways to Enhance Usability, Influence Perception, Increase Appeal, Make Better Design Decisions, and Tech Through Design [25 Additional Design Principles]*: Rockport publishers.
- Liew, C., & Foo, S. (1999). *Derivation of interaction environment and information object properties for enhanced integrated access and value-adding to electronic documents*. Paper presented at the Aslib Proceedings.
- Lin, X. (1997). Map displays for information retrieval. *JASIS*, 48(1), 40-54.
- Lin, X., Soergel, D., & Marchionini, G. (1991). *A self-organizing semantic map for information retrieval*. Paper presented at the Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval.
- Lin, Y., Ahn, J., Brusilovsky, P., He, D., & Real, W. (2010). Imagesieve: Exploratory search of museum archives with named entity-based faceted browsing. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-10. doi: 10.1002/meet.14504701217
- Lindberg, D., Humphreys, B., & McCray, A. (1993). The Unified Medical Language System. *Methods of information in medicine*, 32(4), 281.
- Livnat, Y., Agutter, J., Moon, S., Erbacher, R., & Foresti, S. (2005). *A visualization paradigm for network intrusion detection*. Paper presented at the Information Assurance Workshop, 2005. IAW'05. Proceedings from the Sixth Annual IEEE SMC.
- Longmire, W. (2000). A primer on learning objects. *Learning Circuits*, 1(3).
- Lytras, M., & Sicilia, M. (2007). Where is the value in metadata? *International Journal of Metadata, Semantics and Ontologies*, 2(4), 235-241.

- Mackinlay, J. (1991). Search architectures for the automatic design of graphical presentations. *ACM Press, New York*, 281-292.
- Mackinlay, J., Hanrahan, P., & Stolte, C. (2007). Show me: Automatic presentation for visual analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1137-1144.
- Mangold, C. (2007). A survey and classification of semantic search approaches. *International Journal of Metadata, Semantics and Ontologies*, 2(1), 23-34.
- Marchionini, G. (1997). *Information seeking in electronic environments* (Vol. 9): Cambridge University Press.
- Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4), 41-46.
- Marchionini, G. (2008). Human-information interaction research and development. *Library & Information Science Research*, 30(3), 165-174. doi: 10.1016/j.lisr.2008.07.001
- Marchionini, G., & White, R. (2010). Find what you need, understand what you find. *International Journal of Human Computer Interaction*, 23(3), 205-237.
- Margaryan, A., & Littlejohn, A. (2008). *Repositories and communities at cross-purposes: issues in sharing and reuse of digital learning resources*. 24, Blackwell Publishing Ltd. Retrieved from <http://dx.doi.org/10.1111/j.1365-2729.2007.00267.x> (4)
- Martín, D., Sánchez, S., Gaona, P., & Marianos, N. (2013). *Applying visualization techniques to develop interfaces for educational repositories: the case of Organic.Lingua and VOA3R*. Paper presented at the Proceedings of the Learning Innovations and Quality: The Future of Digital Resources, Rome.
- McGreal, R. (2008). A typology of learning object repositories. *Handbook on Information Technologies for Education and Training*, 5-28.
- McKay, D., Shukla, P., Hunt, R., & Cunningham, S. (2004). Enhanced browsing in digital libraries: three new approaches to browsing in Greenstone. *International Journal on Digital Libraries*, 4(4), 283-297.
- Merčun, T., & Žumer, M. (2009). Visualizing FRBR. *Libraries in the Digital Age, LIDA 2009, 25-30 May, Dubrovnik and Zadar Croatia*, 209-215.
- Merčun, T., & Žumer, M. (2010). *Visualizing for explorations and discovery*. Paper presented at the Proc. of the conf. on Libraries in the Digital Age, Zadar, Croatia.
- Merčun, T., Žumer, M., & Aalberg, T. (2012). FrbrVis: An Information Visualization Approach to Presenting FRBR Work Families Theory and Practice of Digital Libraries. In P. Zaphiris, G. Buchanan, E. Rasmussen & F. Loizides (Eds.), *Theory and Practice of Digital Libraries* (Vol. 7489, pp. 504-507): Springer Berlin / Heidelberg.
- Miles, A., Matthews, B., Wilson, M., & Brickley, D. (2005). *SKOS core: simple knowledge organisation for the web*. Paper presented at the International Conference on Dublin Core and Metadata Applications "Metadata Vocabularies in Practice", 12 - 15 September 2005, Leganés - Madrid, Spain.
- Mizoguchi, R., Vanwelkenhuysen, J., & Ikeda, M. (1995). Task ontology for reuse of problem solving knowledge. *Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing*, 46-57.
- Moen, W., Stewart, E., & McClure, C. (1997). The role of content analysis in evaluating metadata for the us government information locator service (gils): results from an exploratory study. *GILSMDContentAnalysis.htm*.
- Mora, M. (2004). Interacción en interfaces de recuperación de información: conceptos, metáforas y visualización. *Gijón: Trea*.
- Morante, M. (2003). Usability guidelines for taxonomy development. *Montague Institute Review* <http://www.montague.com/abstracts/usability.html> (Acceso: 2 de Mayo, 2014).
- Muelder, C., & Ma, K. (2008). *A treemap based method for rapid layout of large graphs*. Paper presented at the Visualization Symposium, 2008. PacificVIS '08. IEEE Pacific, Tokio.

- Muñoz, J., Calvillo, E., Ochoa, C., Santaolaya, R., & Álvarez, F. (2010). Use of Agents to Realize a Federated Searching of Learning Objects. *Trends in Practical Applications of Agents and Multiagent Systems*, 1-8.
- Nakayama, K., Hara, T., & Nishio, S. (2007, 21-23 May 2007). *A Thesaurus Construction Method from Large Scale Web Dictionaries*. Paper presented at the Advanced Information Networking and Applications, 2007. AINA '07. 21st International Conference on.
- Nash, S. (2005). *Learning objects, learning object repositories, and learning theory: Preliminary best practices for online courses*. 1. (1)
- Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. (1991). Enabling technology for knowledge sharing. *AI magazine*, 12(3), 36.
- Nicholas, D., Huntington, P., & Jamali, H. (2007). The use, users, and role of abstracts in the digital scholarly environment. *The journal of academic librarianship*, 33(4), 446-453.
- Nielsen, J. (1994a). *Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier*. Academic Press, Inc.
- Nielsen, J. (1994b). *Usability engineering*. Morgan Kaufmann Publishers Inc., San Francisco, California, USA.
- Nielsen, J. (2003). *Usability 101: Introduction to usability*. 25.
- Nielsen, J., Snyder, C., Molich, R., & Farrell, S. (2001). *E-commerce user experience*. Nielsen Norman Group.
- Noik, E. (1993). Exploring large hyperdocuments: fisheye views of nested networks. *Proceedings of the fifth ACM conference on Hypertext*, 192-205.
- Norman, D. (2005). Human-centered design considered harmful. *Interactions*, 12(4), 14-19.
- Norman, D. (2008). Simplicity is not the answer. *Interactions*, 15(5), 45-46.
- Noshairwan, W., Qadir, M., & Fahad, M. (2007). Sufficient Knowledge Omission Error and Redundant Disjoint Relation in Ontology *Advances in Intelligent Web Mastering* (pp. 260-265): Springer.
- Ochoa, X., Cardinaels, K., Meire, M., & Duval, E. (2005). *Frameworks for the automatic indexation of learning management systems content into learning object repositories*. Paper presented at the Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005.
- Ochoa, X., & Duval, E. (2006). *Quality Metrics for learning object Metadata*. Paper presented at the Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006, 1 March 2006, Orlando, Florida, USA.
- Ochoa, X., & Duval, E. (2008). Relevance ranking metrics for learning objects. *Learning Technologies, IEEE Transactions on*, 1(1), 34-48.
- Ochoa, X., & Duval, E. (2009). Quantitative analysis of learning object repositories. *Learning Technologies, IEEE Transactions on*, 2(3), 226-238.
- Pancheshnikov, Y. (2007). Integrating print and digital resources in library collections. *Library Collections, Acquisitions, and Technical Services*, 31(2), 111-112.
- Park, J. (2005). Semantic interoperability across digital image collections: A pilot study on metadata mapping. *Proceedings Canadian Association for Information Science (CAIS)*.
- Park, J. (2009). Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & Classification Quarterly*, 47(3-4), 213-228.
- Pastor, J. (2009). Diseño de un sistema colaborativo para la creación y gestión de tesauros en Internet basado en SKOS. *PhD thesis*.
- Perallos, A. (2007). *Metodología ágil y adaptable al contexto para la evaluación integral y sistemática de la calidad de los sitios web*. Universidad de Deusto, Bilbao.
- Petrelli, D. (2008). On the role of user-centred evaluation in the advancement of interactive information retrieval. *Information Processing & Management*, 44(1), 22-38. doi: 10.1016/j.ipm.2007.01.024

- Pirolli, P., Card, S., & Van Der Wege, M. (2001). *Visual information foraging in a focus+ context visualization*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Plaisant, C., Grosjean, J., & Bederson, B. (2002). *Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation*. Paper presented at the INFOVIS 2002. IEEE Symposium on Information Visualization, 2002., USA.
- Polsani, P. (2006). Use and abuse of reusable learning objects. *Journal of Digital information*, 3(4).
- Polson, P., & Lewis, C. (1990). Theory-based design for easily learned interfaces. *Human-Computer Interaction*, 5(2-3), 191-220.
- Poole, A., & Ball, L. (2006). Eye tracking in HCI and usability research. *Encyclopedia of Human-Computer Interaction*, C. Ghaoui (ed.).
- Prabha, C. (2007). Shifting from print to electronic journals in ARL university libraries. *Serials Review*, 33(1), 4-13.
- Pullan, M., Watson, M., Kennedy, J., Raguenaud, C., & Hyam, R. (2000). The Prometheus Taxonomic Model: a practical approach to representing multiple classifications. *Taxon*, 55-75.
- Qing, F., & Ruhua, H. (2008). *Evaluating the usability of discipline repositories*. Paper presented at the IT in Medicine and Education, 2008. ITME 2008. IEEE International Symposium on.
- Reddy, M., & Wang, R. (1995). Estimating data accuracy in a federated database environment. In S. Bhalla (Ed.), *Information Systems and Data Management* (Vol. 1006, pp. 115-134): Springer Berlin Heidelberg.
- Rehak, D., & Mason, R. (2003). Keeping the learning in learning objects. *Reusing online resources: A sustainable approach to e-learning*, 20-34.
- Reilly, D., & Inkpen, K. (2007). *White rooms and morphing don't mix: setting and the evaluation of visualization techniques*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Roberts, J. (1995). Faculty knowledge about library services at the University of the West Indies. *New Library World*, 96(2), 14-22.
- Robertson, G., Mackinlay, J., & Card, S. (1991). *Cone trees: animated 3D visualizations of hierarchical information*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology.
- Rosch, J. L., & Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, technology & work*, 15(3), 313-327.
- Ruecker, S. (2003). *Affordances of prospect for academic users of interpretively-tagged text collections*. University of Alberta.
- Ruecker, S., Radzikowska, M., & Sinclair, S. (2011). *Visual interface design for digital cultural heritage: A guide to rich-prospect browsing*: Ashgate Publishing, Ltd.
- Russell, S., Norvig, P., & Davis, E. (2010). *Artificial intelligence: a modern approach* (Vol. 2): Prentice hall Englewood Cliffs.
- Sánchez, S., & Sicilia, M. (2005). Normative specifications of learning objects and learning processes: towards higher levels of automation in standardized e-learning. *International Journal of Instructional Technology and Distance Learning*, 2(3), 3-12.
- Santos, J., Ochoa, X., Parra, G., & Duval, E. (2011). La Experiencia de ARIADNE: Creando una Red de Reutilización de Objetos de Aprendizaje a través de Estándares y Especificaciones. *IEEE-RITA*, 6(3), 112-117.
- Sanz, J., Dodero, J., & Sánchez, S. (2009). A preliminary analysis of software engineering metrics-based criteria for the evaluation of learning objects reusability. *International Journal of Emerging Technologies in Learning (iJET)*, 4(2009), 30-34.

- Sanz, J., Dodero, J., & Sánchez, S. (2011). Metrics-based evaluation of learning object reusability. *Software Quality Journal*, 19(1), 121-140.
- Schaffert, S., & Geser, G. (2008). Open educational resources and practices. *eLearning Papers*, 7.
- Schulz, H., Hadlak, S., & Schumann, H. (2011). Point-based visualization for large hierarchies. *Visualization and Computer Graphics, IEEE Transactions on*, 17(5), 598-611.
- Sebrechts, M., Cugini, J., Laskowski, S., Vasilakis, J., & Miller, M. (1999). *Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces*. Paper presented at the Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval.
- Shiri, A. (2008). Metadata-enhanced visual interfaces to digital libraries. *Journal of Information Science*, 34(6), 763-775.
- Shiri, A., & Revie, C. (2005). Usability and user perceptions of a thesaurus-enhanced search interface. *Journal of documentation*, 61(5), 640-656.
- Shiri, A., Revie, C., & Chowdhury, G. (2002). Thesaurus-enhanced search interfaces. *Journal of Information Science*, 28(2), 111-122.
- Shiri, A., Ruecker, S., Doll, L., Bouchard, M., & Fiorentino, C. (2011). An evaluation of thesaurus-enhanced visual interfaces for multilingual digital libraries *Research and Advanced Technology for Digital Libraries* (pp. 236-243): Springer.
- Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG)*, 11(1), 92-99.
- Shneiderman, B. (2003). *Designing the user interface*: Pearson Education India.
- Shneiderman, B., Feldman, D., Rose, A., & Grau, X. F. (2000). *Visualizing digital library search results with categorical and hierarchical axes*. Paper presented at the Proceedings of the fifth ACM conference on Digital libraries, Rutgersburg.
- Shneiderman, B., & Johnson, B. (1991). "Tree-maps: A space-Filling Approach to the Visualization of Hierarchical Information Structures." *Proceedings of IEEE Information Visualization* pp 175-282.
- Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface 4 th edition*. ed: Pearson Addison Wesley, USA.
- Shneiderman, B., & Plaisant, C. (2006). *Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies*. Paper presented at the Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization.
- Shreeves, S., Knutson, E., Stvilia, B., Palmer, C., Twidale, M., & Cole, T. (2005). Is 'quality' metadata 'shareable' metadata? The implications of local metadata practices for federated collections. . In *Proceedings of the Association of College and Research Libraries (ACRL) 12th National Conference*. Minneapolis, MN.
- Shuler, J. (2007). Public Policies and Academic Libraries - The Shape of the Next Digital Divide. *The journal of academic librarianship*, 33(1), 141-143.
- Shuling, W. (2007). Investigation and analysis of current use of electronic resources in university libraries. *Library Management*, 28(1/2), 72-88.
- Sicilia, M., & Garcia, E. (2003). On the concepts of usability and reusability of learning objects. *The International Review of Research in Open and Distance Learning*, 4(2).
- Sicilia, M., Sánchez, S., & Benito, M. (2006). Estado de la Cuestión de los Objetos y Diseños para el Aprendizaje y su uso: Technical report, REDAOPA Red Temática de Actividades y Objetos para el Aprendizaje.
- Sifer, M. (2003). *Exploring web site log data with a multi-classification interface*. Paper presented at the Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on.

- Simkin, D., & Hastie, R. (1987). *An information-processing analysis of graph perception*. 82. (398)
- Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., & Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27(3), 431-432.
- Soergel, D. (1995). The art and architecture thesaurus (AAT): A critical appraisal. *Visual Resources*, 10(4), 369-400.
- Soler, C., & Gil, I. (2010). Possibilities and limitations of thesauri in comparison with other systems of knowledge organization: folksonomies, taxonomies and ontologies. *Rev. Interam. Bibliot.*, 361-377.
- Soto, C., Gordo, E., & Sánchez, S. (2007). Semantic learning object repositories. *International Journal of Continuing Engineering Education and Life Long Learning*, 17(6), 432-446.
- Spence, I. (2005). No humble pie: The origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics*, 30(4), 353-368.
- Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology*, 5(1), 61-77.
- Stace, C. (1991). *Plant taxonomy and biosystematics* (2nd edition ed.). UK: Cambridge University Press.
- Stafford, A., Shiri, A., Ruecker, S., Bouchard, M., Mehta, P., Anvik, K., & Rossello, X. (2008). Searchling: user-centered evaluation of a visual thesaurus-enhanced interface for bilingual digital libraries *Research and Advanced Technology for Digital Libraries* (pp. 117-121): Springer.
- Stasko, J., Catrambone, R., Guzdial, M., & McDonald, K. (2000). An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies*, 53(5), 663-694.
- Stawniak, M. (2012). Searching Content Related by Semantics, Space and Time *Interactive 3D Multimedia Content* (pp. 223-251): Springer.
- Stefaner, M., Dalla Vecchia, E., Condotta, M., Wolpers, M., Specht, M., Apelt, S., & Duval, E. (2007). MACE—enriching architectural learning objects for experience multiplication. *Creating New Learning Experiences on a Global Scale*, 4753, 322-336.
- Stefaner, M., & Muller, B. (2007). *Elastic lists for facet browsers*. Paper presented at the Database and Expert Systems Applications, 2007. DEXA'07. 18th International Workshop on.
- Stefaner, M., Wolpers, M., Memmel, M., Duval, E., Specht, M., Börner, D., Klemke, R. (2009). MACE: Joint Deliverable" Evaluation of the MACE system".
- Stojanovic, L., Staab, S., & Studer, R. (2001). *eLearning based on the Semantic Web*. Paper presented at the WebNet2001-World Conference on the WWW and Internet, Orlando, Florida, USA.
- Strauch, C., Sites, U.-L. S., & Kriha, W. (2011). NoSQL databases. URL: <http://www.christof-strauch.de/nosql dbs.pdf> (Acceso: 2 de Mayo 2014).
- Strong, D., Lee, Y., & Wang, R. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103-110. doi: 10.1145/253769.253804
- Stuckenschmidt, H., Vdovjak, R., Houben, G., & Broekstra, J. (2004). *Index structures and algorithms for querying distributed RDF repositories*. Paper presented at the Proceedings of the 13th international conference on World Wide Web.
- Sutcliffe, A., Ennis, M., & Hu, J. (2000). Evaluating the effectiveness of visual user interfaces for information retrieval. *International Journal of Human-Computer Studies*, 53(5), 741-763.
- Swan, A., & Carr, L. (2008). Institutions, their repositories and the web. *Serials Review*, 34(1), 31-35.
- Swan, R., & Allan, J. (1998). *Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval systems*. Paper presented at the Proceedings of the 21st annual

- international ACM SIGIR conference on Research and development in information retrieval.
- Tadapak, P., Suebchua, T., & Rungsawang, A. (2010). *A machine learning based language specific web site crawler*. Paper presented at the Network-Based Information Systems (NBIS), 2010 13th International Conference on.
- Tenopir, C. (2003). *Use and users of electronic library resources: An overview and analysis of recent research studies*. Paper presented at the Council on Library and Information Resources, Washington (2003).
- Ternier, S., Verbert, K., Parra, G., Vandeputte, B., Klerkx, J., Duval, E., Ochoa, X. (2009). The Ariadne Infrastructure for Managing and Storing Metadata. *Internet Computing, IEEE*, 13(4), 18-25. doi: 10.1109/mic.2009.90
- Tilkov, S., & Vinoski, S. (2010). Node.js: Using JavaScript to Build High-Performance Network Programs. *IEEE Internet Computing*, 14(6).
- Tractinsky, N., Katz, A., & Ikar, D. (2000). What is beautiful is usable. *Interacting with computers*, 13(2), 127-145.
- Tripathy, A., & Patra, P. (2008). *A Web Mining Architectural Model of Distributed Crawler for Internet Searches Using PageRank Algorithm*. Paper presented at the Asia-Pacific Services Computing Conference, 2008. APSCC '08. IEEE.
- Trust, J. (1988). Art & Architecture Thesaurus (AAT). available online at: http://www.getty.edu/research/conducting_research/vocabularies/aat/about.html, (Acceso: 2 de Mayo 2014).
- Tsakonas, G., & Papatheodorou, C. (2006). Analysing and evaluating usefulness and usability in electronic information services. *Journal of Information Science*, 32(5), 400-419.
- Tsakonas, G., & Papatheodorou, C. (2007). *Critical constructs of digital library interaction*. Paper presented at the 11th Panhellenic Conference on Informatics, 18-20 May, Πάτρα (GR).
- Tsakonas, G., & Papatheodorou, C. (2008). Exploring usefulness and usability in the evaluation of open access digital libraries. *Information processing & management*, 44(3), 1234-1250.
- Tudhope, D., Binding, C., Blocks, D., & Cunliffe, D. (2006). Query expansion via conceptual distance in thesaurus indexed collections. *Journal of documentation*, 62(4), 509-533.
- Tunkelang, D. (2009). Faceted search. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1), 1-80.
- Uddin, M., & Janecek, P. (2007). Performance and usability testing of multidimensional taxonomy in web site search and navigation. *Performance measurement and metrics*, 8(1), 18-33.
- ur Rehman, S., & Ramzy, V. (2004). Awareness and use of electronic information resources at the health sciences center of Kuwait University. *Library Review*, 53(3), 150-156.
- Valente, A. (2005). Types and roles of legal ontologies *Law and the semantic web* (pp. 65-76): Springer.
- van Wijk, J., & van de Wetering, H. (1999). *Cushion Treemaps: Visualization of Hierarchical Information*. Paper presented at the (INFOVIS99). IEEE Symposium on Information Visualization, San Francisco, California.
- Veerasamy, A., & Belkin, N. (1996). *Evaluation of a tool for visualization of information retrieval results*. Paper presented at the Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval.
- Vessey, I., & Galletta, D. (1991). *Cognitive fit: An empirical study of information acquisition*. 2. (1)
- von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J., Fekete, J., & Fellner, D. (2011). *Visual Analysis of Large Graphs: State-of-the-Art and Future*

- Research Challenges*. 30, Blackwell Publishing Ltd. Retrieved from <http://dx.doi.org/10.1111/j.1467-8659.2011.01898.x> (6)
- Vora, P. (1998). Design/methods & tools: Designing for the web: a survey. *Interactions*, 5(3), 13-30.
- W3C, H. (2013). The canvas element (Editor's Draft 24 July 2013). <http://dev.w3.org/html5/spec/Overview.html#the-canvas-element> (Acceso: 2 de Mayo, 2014).
- Wang, R., & Strong, D. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 5-33.
- Wang, Z., Chaudhry, A., & Khoo, C. (2008). Using classification schemes and thesauri to build an organizational taxonomy for organizing content and aiding navigation. *Journal of documentation*, 64(6), 842-876.
- Weinberg, B. (1995). Art & Architecture Thesaurus ; and Guide to Indexing and Cataloging with the Art & Architecture Thesaurus. *Journal of the American Society for Information Science*, 46(2), 152-160.
- Westerman, S., & Cribbin, T. (2000). Mapping semantic information in virtual space: dimensions, variance and individual differences. *International Journal of Human-Computer Studies*, 53(5), 765-787.
- White, R., Kules, B., & Drucker, S. (2006). Supporting exploratory search, introduction, special issue, communications of the ACM. *Communications of the ACM*, 49(4), 36-39.
- White, R., & Roth, R. (2009). Exploratory Search: Beyond the Query-Response Paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1), 1-98. doi: 10.2200/s00174ed1v01y200901icr003
- Whitehall, T. (1992). Quality in library and information service: a review. *Library Management*, 13(5), 23-35.
- Whitehall, T. (1995). Value in library and information management: a review. *Library Management*, 16(4), 3-11.
- Wiley, D. (2002a). Connecting learning objects to instructional design theory: A definition, a metaphor and a taxonomy. In *The instructional use of learning objects*. Bloomington, Indiana, 2830(435), 3-24.
- Wiley, D. (2002b). *Instructional use of learning objects*: Agency for Instructional Technology.
- Wiley, D., & Hilton, J. (2009). Openness, dynamic specialization, and the disaggregated future of higher education. *The International Review of Research in Open and Distance Learning*, 10(5).
- Wilkinson, L., & Wills, G. (2005). *The grammar of graphics*: Springer Science+ Business Media.
- Wise, J. (1999). The ecological approach to text visualization. *Journal of the American Society for Information Science*, 50(13), 1224-1233.
- Wittenburg, K., Das, D., Hill, W., & Stead, L. (1995). *Group asynchronous browsing on the World Wide Web*. Paper presented at the 4th International WWW Conference, Boston (MA).
- Wu, M., Fuller, M., & Wilkinson, R. (2001). Using clustering and classification approaches in interactive retrieval. *Information processing & management*, 37(3), 459-484.
- Wylie, B., & Baumes, J. (2009). *A unified toolkit for information and scientific visualization*. Paper presented at the IS&T/SPIE Electronic Imaging.
- Yang, C., Chen, H., & Hong, K. (1999). *Visualization tools for self-organizing maps*. Paper presented at the Proceedings of the fourth ACM conference on Digital libraries, Berkeley, California, USA.
- Ytow, N., Morse, D., & Roberts, D. (2001). Nomencurator: a nomenclatural history model to handle multiple taxonomic views. *Biological journal of the Linnean Society*, 73(1), 81-98.

- Zhang, J. (2001). TOFIR: A tool of facilitating information retrieval – introduce a visual retrieval model. *Information Processing & Management*, 37(4), 639-657. doi: 10.1016/s0306-4573(00)00042-x
- Zhang, J., & Korfhage, R. (1999). DARE: distance and angle retrieval environment: a tale of the two measures. *Journal of the American Society for Information Science*, 50, 9.
- Zhang, S., Shen, W., & Ghenniwa, H. (2004). A review of Internet-based product information sharing and visualization. *Computers in Industry*, 54(1), 1-15.
- Zhong, Y., Luo, Y., Pramanik, S., & Beaman, J. (1999). HICLAS: a taxonomic database system for displaying and comparing biological classification and phylogenetic trees. *Bioinformatics*, 15(2), 149-156.